International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/

(Research Article)

Measuring student skills in explaining socio-scientific chemical phenomena using rasch model

Rahmat Utina *, Lukman A.R Laliyo and Rustam I. Husain

Educational Technology Department, Universitas Negeri Gorontalo, Jl. Jenderal Sudirman No. 06 Gorontalo City 96128, Indonesia

International Journal of Science and Research Archive, 2023, 09(02), 488–495

Publication history: Received on 11 June 2023; revised on 20 July 2023; accepted on 22 July 2023

Article DOI: https://doi.org/10.30574/ijsra.2023.9.2.0566

Abstract

This non-experimental research applied a descriptive-quantitative approach to measure student skills in explaining socio-scientific chemical phenomena. The data collection technique employed was a tier multiple choice test composed of three tier questions: Q1, Q2, and Q3. The population comprised 703 eleventh graders in Gorontalo Province. The analysis exerted was Rasch modelling. The results skills in explaining socio-scientific chemical phenomena.

Keywords: Rasch Modelling; Explaining Skills; Socio-Scientific Chemical Phenomena

1. Introduction

Information technology advance during the 4.0 Revolution greatly affects education systems in this new-normal condition. The advance, along with communication advance, can alter the teacher-centered-learning principle to the student-centered-learning one (Septantiningtyas, 2018). The Revolution also impacts the educational world in Indonesia, in which the government desires a digitalized education system obliging all educational elements to adjust to the current need (Septiawan et al., 2019). Education during the 4.0 Revolution aims to achieve smart education through augmented and equal distribution of quality education and expanded accessibility and relevance to technology, attaining world-class quality education (Noermanzah & Friantary, 2019). Technology advance during the 4.0 Industrial Revolution creates new challenges in chemistry learning. It changes learning patterns, covering how students learn. The COVID-19 pandemic brings about a basic shift in using technology for learning chemistry. It augments the need for information technology. Meanwhile, chemistry learning challenges are centered to how to provide certain learning that can elevate conceptual understanding of students.

In learning, concepts delivered should be understood, learned, and mastered. In chemistry learning, all materials are interconnected. Ill-conceived previous materials will bring about difficulties in apprehending the following ones (Hikmah, 2017). According to Maghfiroh et al. (2016), students acquired conceptual understanding from learning during a learning activity. Students should understand a knowledge concept to interpret materials in their own language without referencing to books. Basic concepts must be well understood before comprehending other complex ones. Accordingly, conceptual understanding is the basics of learning.

Chemistry assists students to understand phenomena around. Socio-scientific phenomena are chemistry-related topics as regards daily life. Through learning these phenomena, there will be many daily phenomena explorable. Being familiarized with chemistry-related concepts, students will be able to explain them.

* Corresponding author: Rahmat Utina

And yet, chemistry learning is still traditional. In line with Surpi (2017), learning is still traditional, and accordingly, needs to be transformed into a modern as to tackle global challenges.

A challenge in cultivating student skills lays in student habits in which student is not accustomed to environmental problems in formal learning (Orwat et al., 2017). As a result, students are difficult to exploit their knowledge and understanding of science to explain daily problems (Hofstein et al*.,* 2011). It causes students to be unable to relate particular knowledge to their understanding (Laliyo, 2021). Additionally, students are inclined to consider chemistry an unimportant subject as it is not correlated with their life (Aikenhead, 2003). Poor learning will engender more students that cannot understand concepts and explain the learned concepts.

Students must acquire explaining skills. In the scope of skills, explaining skills, as Ubabuddin (2020) argues, are orally presenting information in a structured method to exhibit the relationship between a cause and effect, between a definition and its examples, or between a definition and a related unknown object. Students are considered having good explaining skills once they can understand the delivered concepts.

To the extent to which students can explain a concept can be measured using their skills. A test is one of the effective ways to measure student skills of explaining social chemistry phenomena. It measures student skills of explaining a concept. Measuring student explaining skills enables us to identify to what extent the skills are. As such, measuring student skills of explaining a concept can be used to evaluate learning processes and acts as a solution to problems, giving off better learning. Hence, measuring student skills of explaining a concept is salient.

Another research on the Rasch model indicates student critical thinking skills. Sabekti & Khoirunnisa (2018) put the Rasch model to develop instruments to measure students' critical thinking skills related to chemical bond topics. They found that in designing instruments to measure critical thinking skills in chemical planes, using the Rasch model is promising.

2. Method

This non-experimental research used a descriptive-quantitative method. As conveyed by Sugiyono (2018), descriptive-quantitative research defined a subject or object. In the research, researchers gave no treatment to respondents. The current research aimed to measure student skills of explaining socio-scientific chemistry phenomena. The test classes were several schools in Gorontalo Province with extended demographic aspects of respondents. In so doing, this research could point out different skills of students from some schools. This research was performed in the even semester in the academic year of 2021/2022 at senior high schools in Gorontalo Province. The subjects were senior-high-school students in Gorontalo Province that had been delivered chemistry for eleventh graders in science classes.

The sampling technique was purposive sampling. We determined samples under specific considerations. Associated with ethical considerations, we asked informed consent from respondents to be willing to participate in this research before filling out test instruments. Respondents were senior-high-school students randomly taken from ten schools in Gorontalo. They were 703 students aged 15-22 years old and had been delivered lessons for eleventh graders. We applied a test as a data collection technique, directly implemented on students in a face-to-face meeting. The test was in the form of a tier multiple choice diagnostic test given to 703 senior-high-school students to measure their skills in explaining socio-scientific chemistry phenomena. We supervised the text implementation to collect data. Data were collected for two months. In the test, undertaken once, students were instructed to write their responses on answer sheets and given time one hour and 30 minutes.

2.1. Instrument and Procedure

The instrument was a diagnostic test. A diagnostic test aimed to examine student strength and weaknesses when learning a material. The results would serve as the consideration for giving required follows-up (Suteno et al., 2021). The instrument was a tier multiple-choice test to measure student skills in explaining socio-scientific chemistry phenomena. The test was made up of three tier questions, i.e., Q1, Q2, and Q3.

The instrument was validated by an expert validated and tested for its reliability. Once declared as valid and reliable, it could be exerted to measure student skills in explaining socio-scientific chemistry phenomena. In designing this instrument, we referred to Wilson (2005) with his four seminal facets: the construct map, item, assessment, and measurement model.

2.2. Data Analysis

Data were analyzed using the Rasch model. The Rasch model measurement aimed to induce a measurement scale at the same interval to provide accurate information about test participants and the quality of the solved items (Sumintono & Wihdiarso, 2015:15). The Rasch model could predict missing data. Applying the model allowed our statistical analysis results to result in the best data from the missing data (Sumintono & Wihdiarso, 2015:46). The Rasch model output formed calibration quality on three objects: measurement scale, respondent (person), and item. The data analysis, as posited by Laliyo (2021:36-37), was carried out in the seven following stages.

- First: evaluating scores required from student responses to all questions (Q1, Q2, and Q3) for each item into data. The resulted data were "uniform" polytomy data in the form of ranks or ratings at maximum and minimum scores of three or zero, respectively.
- Second: making sure that the polytomy data were in an EXCEL format based on sex.
- Third: converting raw scores into the same interval sizes using the WINSTEP 3.73 software.
- Fourth: estimating the effectiveness of the exploited instrument by validity and reliability of each person and item.
- Fifth: estimating item validity by implementing item statistics test: misfit order.

3. Result and Discussion

3.1. Person and Item Reliability

Measurement instrument effectiveness could be based on person and item reliability. Measurement aimed to identify the degree to which measurement brought about consistent information that could disclose latent traits od unidimensional characteristics of the measured variable (Sumintono & Widhiarso, 2015).

The person and item reliability analysis test results are presented in Table 1. From Table 1, the total number of data was 21090, bringing on a Chi-square of 46596.41 at the degree of freedom of 46596.41 (p = 0.0000), exhibiting that the measurement was good and significant. The person reliability value was 0.83, equivalent to the person separation index value of 2.25. It indicated that student response consistency to the test was good. The KR-20 value (Cronbach's alpha) of 0.84 pointed out a good interaction between students and items. Accordingly, the actual data fulfilled the Rasch model requirements, and a further analysis could be conducted. It showed a strong correlation between student responses and items, in which student knowledge was fragmented (Adams & Wieman, 2011) and as such, measurable. Hence, this instrument was reliable to differ student skills in explaining socio-scientific chemistry phenomena well.

Table 1 Person and Item Reliability

	Person (703)	**Item (30)**
Mean	0.37	0.00
Standard error	0.02	0.08
Standard Deviation (SD)	0.51	0.43
Reliability	0.83	0.99
INFIT MNSQ	1.02	1.00
OUFIT MNSQ	1.02	1.02
INFIT ZSTD	-0.10	-0.20
OUTFIT ZSTD	-0.10	0.30
Point Raw Score to measure correlation	0.99	-1.00
Separation Index (reliability)	2.25	9.68
Cronbach Alpha (KR-20) = 0.84		
Data Points: 21090		
Chi-Square: 46596.41		
df: 46596.41 (p = 0.0000)		

Relevant information was urgent for educational researchers and practitioners to prepare following-up plans and develop student skills (Wei et al., 2012). The item separation index value of 9.68 was high. That was, all instruments were good, especially its instrument reliability value of 0.99. It suggested better item consistency, or items met unidimensional requirements. That was, items could define the measured variable well. This conclusion was confirmed through the acquired item infit and outfit values, that mostly existed within an acceptable range for a multiple-choice test (Bond & Fox, 2015; Herrmann-Abell & Deboer, 2016).

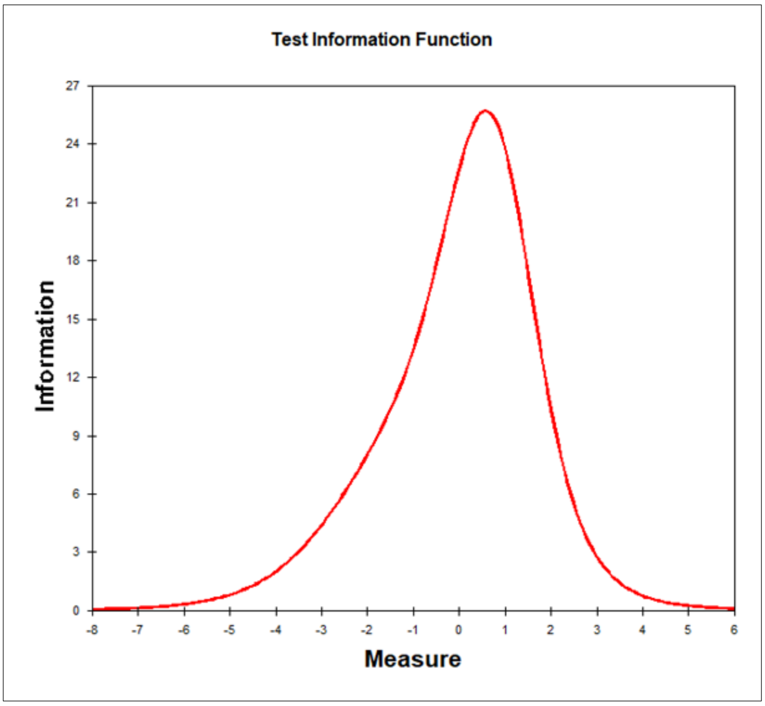

Figure 1 Measurement Information Function

Figure 1 demonstrates a measurement information chart to exhibit measurement reliability. The higher the chart peak, the higher the measurement reliability value. Measurement information was high at a medium student skill level (-4.0 logits to +4.0 logits). It indicated that the skill instrument, in explaining the used socio-scientific chemistry phenomena, could engender optimum information to medium-skilled students. That was, the instrument had good measurement reliability (Misbach & Sumintono, 2014; Sumintono & Widhiarso, 2014; Bond & Fox, 2015).

3.2. Validity

Item validity was measured using the Fit item test to ensure that all items suited the Rasch model. It aimed to investigate if test items measured what should be measured or test validity (Linacre, 2012; Sumintono, 2018). Three criteria applied to observe unsuitable students and items (outliers or misfits) (Linacre, 2012, 2020; Sumintono & Widhiarso, 2015) were the criterion of the accepted Outfit Mean Square (MNSQ) = 0.5 < MNSQ < 1.5, the criterion of the accepted Outfit Z-Standard (ZSTD) = -2.0 < ZSTD < 2.0, and the criterion of the accepted Point Measure Correlation (Pt Mea Corr) = 0.4 < Pt Mea Corr < 0.8. When all criteria were unfulfilled, items were poor and needed a further analysis (Boone et al., 2014). MNSQ outfits and infits were Chi squares sensitive to detect an outlier response pattern. Two types of outlier were correct responses predicted by low-skill students to items with a high difficulty and incorrect responses generated by the carelessness of high-skill students to items with a low difficulty. The ideal MNSQ value was 1.0 logit. The PTMEA CORR value pointed out a correlation between item scores and person measure, whose value should be positive and not close to 0 (Bond & Fox, 2015). Table 2 presents the item statistics analysis results.

In Table 2, we showed that (a) all items met Outfit MNSQ criteria, (b) item-A1, item-A2, item-A3, item-A4, item-B10, item-B15, item-B17, item-B18, item-B23, item-C26, item-C27, item-C28, item-C29, and item-C30 unfulfilled Outfit ZSTD criteria, and (c) no item was valued negative for the PTMEA CORR criteria. Although several items unmet criteria, their quality remained. For example, item-A1, item-A2, item-A3, item-B5, item-B7, item-B8, item-B9, item-B11, item-B12, item-B16, item-B18, and item-C25 unfulfilled PTMEA CORR criteria. And yet, they were still considered valued as there was no PTMEA CORR with a negative value. It suggested that there was no item nt meeting the three criteria or misfits. That was, the measurement instrument had compatible and valid items.

Table 2 Item Statistics: Misfit Order

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
18	774	703	.71	.05	.56	-9.7	.61	-8.0	.38	.37	65.0	52.1	item-B18
25	852	703	.55	.04	.97	-.5	1.07	1.4	.30	.39	40.8	48.0	item-C25
7	901	703	.46	.04	.96	-.9	.99	-.2	.35	.39	46.9	46.0	item-B7
5	903	703	.45	.04	.91	-2.0	1.06	1.2	.14	.39	47.8	46.0	item-B5
20	905	703	.45	.04	.89	-2.4	.91	-1.7	.42	.39	47.9	46.0	item-B20
11	919	703	.42	.04	.92	-1.7	.96	-.7	.25	.40	49.5	44.8	item-B11
9	955	703	.36	.04	.91	-2.0	1.00	.0	.28	.40	44.1	41.9	item-B9
16	970	703	.33	.04	1.07	1.6	1.09	1.7	.33	.41	39.1	40.9	item-B16
19	990	703	.29	.04	1.02	.6	1.06	1.2	.43	.41	35.4	40.8	item-B19
10	994	703	.28	.04	1.11	2.4	1.16	3.2	.44	.41	32.3	39.6	item-B10
27	1009	703	.26	.04	.80	-4.9	.83	-3.9	.44	.41	43.8	39.6	item-C27
29	1027	703	.23	.04	.83	-4.3	.86	-3.2	.54	.41	43.7	38.5	item-C29
22	1038	703	.21	.04	.95	-1.2	.99	-.2	.42	.42	30.6	37.4	item-B22
8	1051	703	.18	.04	.89	-2.7	.91	-2.0	.36	.42	39.4	37.3	item-B8
14	1064	703	.16	.04	.95	-1.1	.96	-.8	.41	.42	32.7	36.0	item-B14
13	1077	703	.14	.04	1.04	1.0	1.07	1.5	.45	.42	34.4	36.0	item-B13
6	1109	703	.08	.04	.97	-.8	.98	-.3	.54	.43	36.4	34.0	item-B6
30	1113	703	.08	.04	.88	-3.1	.88	-2.8	.55	.43	36.6	34.0	item-C30
3	1223	703	-.11	.04	1.17	4.3	1.20	4.5	.39	.44	25.9	30.2	item-A3
21	1288	703	-.22	.04	.89	-2.9	.92	-1.9	.51	.44	32.3	29.6	item-B21
1	1314	703	-.27	.04	1.23	5.4	1.24	5.3	.35	.45	20.5	29.3	item-A1
4	1347	703	-.33	.04	.85	-4.1	.87	-3.1	.53	.45	31.6	28.9	item-A4
2	1389	703	-.40	.04	1.43	9.6	1.46	9.2	.37	.45	28.3	29.8	item-A2
24	1390	703	-.40	.04	1.05	1.3	1.05	1.2	.57	.45	23.9	29.8	item-B24
28	1416	703	-.45	.04	.91	-2.3	.88	-2.9	.58	.45	35.7	30.5	item-C28
26	1435	703	-.48	.04	1.16	3.8	1.16	3.4	.50	.45	24.5	31.3	item-C26
17	1485	703	-.57	.04	1.14	3.1	1.12	2.4	.45	.45	37.1	33.6	item-B17
23	1560	703	-.72	.04	1.17	3.6	1.14	2.6	.50	.45	36.3	38.4	item-B23
12	1561	703	-.72	.04	1.01	.2	.99	-.1	.37	.45	39.1	38.5	item-B12

3.3. Wright Map

The map was made to measure consistency in item difficulty levels and student skill levels. A higher item difficulty level gave off a higher student skill level. Figure 2 demonstrates the information from Wright Map: Person-Map-Item.

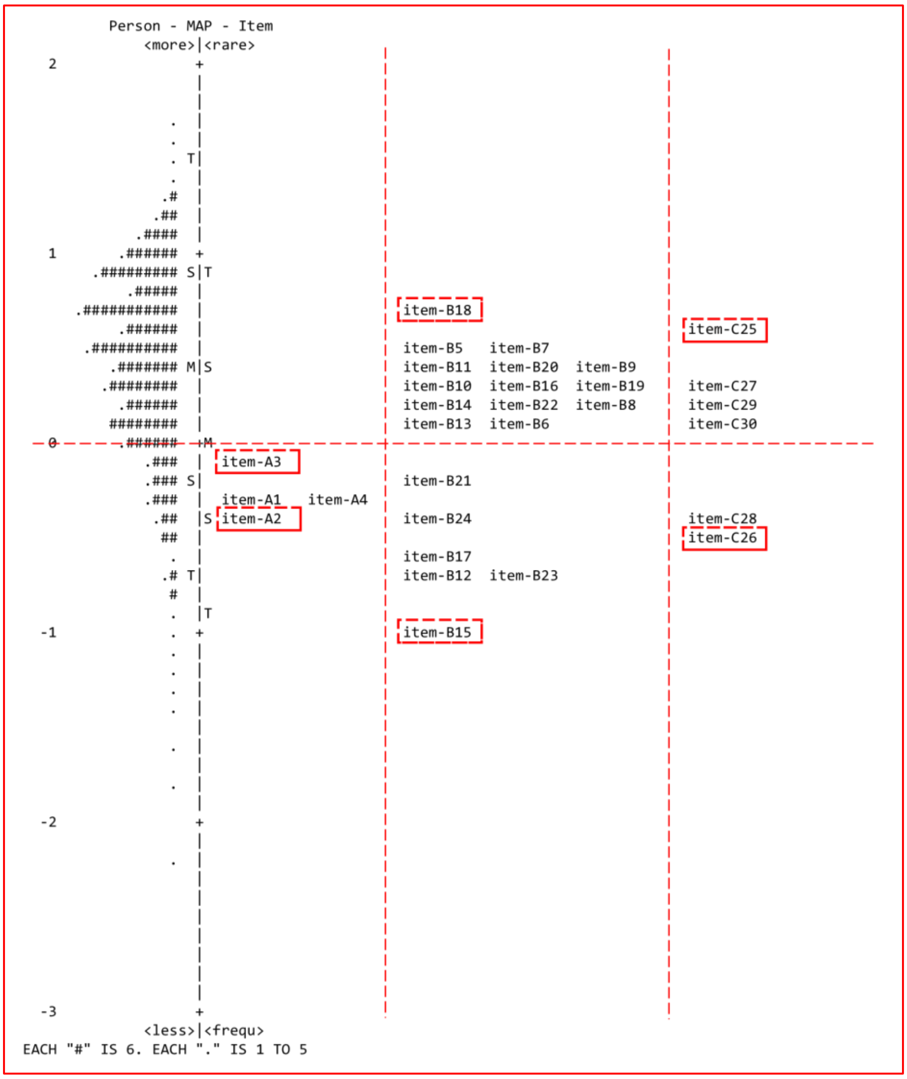

Figure 2 Wright Map: Person (N = 703) and Item (N = 30)

From Figure 2, the Wright Map constituted a chart representing item difficulty levels and student skill levels. This map is a product of the empirical analysis of student responses to all items. All instrument items could cover most student skills. Some students were identified for having a high skill (> +2.0 logits), and some others had a very low skill (< -2.0 logits). Most of the item difficulty levels existed in intervals (-1.0 logit to +2.0 logits), even several items had the same item difficulty level. Item-B18 had the highest difficulty level (+0.71 logits), whereas item-B15 had the lowest one (-0.96 logits).

In terms of item size differences, some cases could be explicated as follows: first, item-B18 (+0.71 logits) > item-B7 (+0.46 logits) > item-B5 (+0.45 logits) > item-B20 (+0.45 logits) > item-B11 (+0.42 logits) > item-B9 (+0.36 logits) > item-B16 (+0.33 logits) > item-B19 (+0.29 logits) > item-B10 (+0.28 logits) > item-B22 (+0.21 logits) > item-B8 (+0.18 logits) > item-B14 (+0.16 logits) > item-B13 (+0.14 logits) > item-B6 (+0.08 logits) > item-B21 (-0.22 logits) > item-B24 (-0.4 logits) > item-B17 (-0.57 logits) > item-B12 (-0.72 logits) > item-B23 (-0.72 logits) > item-B15 (-0.96 logits), in which item-B18 (+0.71 logits) had the highest difficulty level. It demonstrated that students found more difficulties in responding to eleventh-grader materials vis a vis the hydrolysis phenomenon. It exhibited that a few numbers of students could illuminate and understand the knowledge, while others could not provide evidence and solely memorized.

Second, item-C25 (+0.55 logits) > item-C27 (+0.26 logits) > item-C29 (+0.23 logits) > item-C30 (+0.08 logits) > item-C28 (-0.45 logits) > item-C26 (-0.48 logits), where item-C25 (+0.55 logits) had the highest difficulty level. It indicated that students were difficult to respond to twelfth-grader materials with regard to metal extrapolation phenomena. Only few students could lay out and understand the knowledge, whereas others could not provide evidence and merely memorized.

Third, item-A3 (-0.11 logits) > item-A1 (-0.27 logits) > item-A4 (-0.33 logits) > item-A2 (-0.40 logits) >, whereby item-A3 (-0.11 logits) had the highest difficulty level. It suggested that students found more difficulties in responding to tenth-grader materials with respect to the apple color change phenomenon. Students that could set forth and understand the knowledge came in a few numbers, while others that could not provide evidence and solely hung on memorizing were many.

4. Conclusion

The instrument was considered valid and reliable to measure student skills in explaining social-scientific chemistry phenomena. Measuring student skills in explaining the phenomena using the instrument, we figured out that the majority of students could not shed light on and understand the knowledge contained in tenth, eleventh, and twelfth-grader materials. Other students also could not provide evidence and memorized only.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to disclosed.

Statement of informed consent

Informed consent was obtained from all individual participants included in the study.

References

[1] Aikenhead, G. S. (2003). Chemistry and Physics Instruction: Integration, Ideologies, and Choices. *Chem. Educ. Res. Pract.*, *4*(2), 115–130. https://doi.org/10.1039/b2rp90041f

[2] Bond, T. G. & Fox, C. M. (2015). *Applying the Rasch Model: Fundamental Measurent in the Human Sciences* (3rd ed). Routledge Taylor & Francis Group. https://doi.org/10.1088/1751-8113/44/8.085201

[3] Bond, T. G. & Fox, C. M. (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. *Routledge Taylor & Francis Group* (Third Edit, Vol. 44). New York and London: Routledge Taylor & Francis Group. https://doi.org/10.1088/1751-8113/44/8/085201

[4] Boone, W. J., Yale, M. S., & Staver, J. R. (2014) *Rasch analysis in the human sciences*, *Rasch Analysis in the Human Sciences*. doi: 10.1007/978-94-007-6857-4.

[5] Herrmann-Abell, C. F. & Deboer, G. E. (2016). Using Rasch Modeling and Option Probability Curves to Diagnose Students' Misconceptions. Aera, 1–12.

[6] Hikmah, R. (2017). Penerapan Model Advance Organizer untuk Meningkatkan Kemampuan Pemahaman Siswa. *SAP (Susunan Artikel Pendidikan)*, *1*(3), 271–280. https://doi.org/10.30998/sap.v1i3.1204

[7] Hofstein, A., Eilks, I., & Bybee, R. (2011). Societal issues and their importance for contemporary science education-a pedagogical justification and the state-of-the-art in Israel, Germany, and the USA. *International Journal of Science and Mathematics Education*, *9*(6), 1459–1483. https://doi.org/10.1007/s10763-010-9273-9

[8] Laliyo, L.A.R. (2021). Mendiagnosis Sifat Perubahan Konseptual Siswa: Penerapan Teknik Analisis Stacking dan Racking Rasch Model. Yogyakarta: Deepublish.

[9] Linacre, J. M. (2012). A User's Guide to W I N S T E P S ® M I N I S T E P Rasch-Model Computer Programs Program Manual 3.75.0. https://doi.org/ISBN 0-941938-03-4

[10] Linacre, J. M. (2020). A User's Guide to W I N S T E P S ® M I N I S T E P Rasch-Model Computer Programs Program Manual 4.5.1. https://doi.org/ISBN 0-941938-03-4

[11] Maghfiroh, L., Santosa, S., & Suryadharma, I. B. (2016). Identifikasi Tingkat Pemahaman Konsep Stoikiometri pada Pereaksi Pembatas dalam Jenis-jenis Reaksi Kimia Siswa Kelas X MIA Negeri 4 Malang. *Jurnal Pembelajaran Kimia (J-PEK)*, *01*(2), 32–37. https://doi.org/10.30998/sap.v1i3.1204

[12] Misbach, I. & Sumintono, B. (2014). Pengembangan dan Validasi Instrumen Persepsi Siswa tehadap Karakter Moral Pengembangan dan Validasi Instrumen "Persepsi Siswa tehadap Karakter Moral Guru" di Indonesia dengan Model Rasch 1. *Pengembangan Instrumen Penilaian Karakter Yang Valid, May.*

[13] Noermanzah & Friantary, H. (2019). Development of Competency-Based Poetry Learning Materials for Class X High Schools. *International Journal of Recent Technology and Engineering*, *8*(4), 6631–6638. https://doi.org/10.35940/ijrte.d8855.118419

[14] Orwat, K., Bernard, P., & Migdał-Mikuli, A. (2017). Alternative conceptions of common salt hydrolysis among upper-secondaryschool students. *Journal of Baltic Science Education*, *16*(1), 64–76. https://doi.org/10.33225/jbse/17.16.64

[15] Sabekti, A. W. & Khoirunnisa, F. (2018). Penggunaan Rasch Model untuk Mengembangkan Instrumen Pengukuran Kemampuan Berpikir Kritis Siswa pada Topik Ikatan Kimia. *Jurnal Zarah*, *6*(2), 68–75. https://doi.org/10.31629/zarah.v6i2.724

[16] Septantiningtyas, N. (2018). Pengaruh Pembelajaran Jarak Jauh dengan Aplikasi *Google Class* terhadap Hasil Belajar Mahasiswa. *Edureligia; Jurnal Pendidikan Agama Islam*, *2*(2), 131–135. https://doi.org/10.33650/edureligia.v2i2.714

[17] Septiawan, F., Azizah, N., Gita, P. P., & Khery, Y. (2019). Pentingnya Pembelajaran Mobile, Kearifan Lokal, dan Kepariwisataan. *Bioscientist : Jurnal Ilmiah Biologi*, *7*(2), 146–150. https://doi.org/10.33394/bjib.v7i2.2379

[18] Sugiyono. 2018. Metode Penelitian Pendidikan Kuantitatif, Kualitatif, dan R&D. Bandung: CV. Alfabeta.

[19] Sumintono, B. & Widhiarso, W. (2014). Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial (B. Trim (ed.); Issue November). Trim Komunikata Publishing House. https://www. researchgate.net/publication/268688933%0AAplikasi

[20] Sumintono, B. & Widhiarso, W. (2015). Aplikasi Pemodelan Rasch pada *Assessment* Pendidikan (Issue September). Penerbit Trim Komunikata, Cimahi. https://www.researchgate.net/ publication/282673464%0AAplikasi

[21] Sumintono, B. (2018). Rasch model measurements as tools in assesment for learning. Proceedings of the 1st International Conference on Education Innovation (ICEI 2017), October 2017. https://doi.org/10.2991/icei-17.2018.11

[22] Surpi, N. K. (2017). Hadapi Tantangan Global, Lembaga Pendidikan Hindu Harus Jadi Gurukula Modern. *Jurnal Penjaminan Mutu*, *3*(2), 171. https://doi.org/10.25078/jpm.v3i2.197

[23] Suteno, I. K., Laliyo*, L. A. R., Iyabu, H., & Abdullah, R. (2021). Mengevaluasi Level Pemahaman Konsep Hidrolisis Garam Peserta Didik Menggunakan Tes Diagnostik Pilihan Ganda Empat Tingkat. *Jurnal Pendidikan Sains Indonesia*, *9*(3), 482–497. https://doi.org/10.24815/jpsi.v9i3.20543

[24] Ubabuddin, U. (2020). Pelaksanaan Supervisi Pembelajaran sebagai Upaya Meningkatkan Tugas dan Peran Guru dalam Mengajar. *Nidhomul Haq: Jurnal Manajemen Pendidikan Islam*, *5*(1), 102–118. https://doi.org/10.31538/ndh.v5i1.512

[25] Wilson, M (2005). Constructing measure: an item response modeling approach. *Lawrence Erlbaum Associates, Inc.* https://doi.org/10.4324.9781410611697

[26] 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(3), 335-345. https://doi.org/ 10.1021/ed100852tl