The Ability of High-School Students to Solve STEM-Related Problems

By: Mr. Septian Karyana
Training Officer

Mr. Reza Setiawan
Head of the Programme and Training Division

Dr. Indrawati
Centre Director
SEAQIS

Abstract

STEM education has become an important issue for Southeast Asia in the 21st century. This article investigates high-school students' ability to solve STEM problems through a quantitative study. We asked students to answer 10 STEM questions. Their responses were then reviewed, validated, and statistically analysed using the Rasch model by experts from a college. Some 301 11th-grade students from five elementary schools in West Java participated in the study. The questions tested four levels of student cognitive dimension—understanding, applying, analysing, and evaluating. Our statistical analysis showed that the STEM questions used were reliable with a reliability score of 0.93 and so they were fit or acceptable compared with the normal standard. The construct's validity was analysed using the infit and outfit mean square (MNSQ) values. With that, we determined that the instrument was valid. The statistical analysis showed that the students had a low ability to solve STEM problems, particularly when it came to the cognitive dimension. The study provided an overview of the students' ability to solve STEM problems and so is expected to serve as a reference for teachers when planning and designing their own lessons.

Introduction

The 21st century is characterised by ICT taking a central role in all aspects of people's lives. Technology connects everyone around the world, transcending geographical barriers and creating a borderless community. Technological developments have improved the quality of products and services and, as a result, life as we know it. This, however, requires a better and more competent workforce. The 21st century is marked by scattered information and ever-evolving technologies. Today, the world of science has become increasingly interconnected, resulting in better synergy. Various efforts to improve the quality of education are constantly being exerted, including redesigning curricula, developing new teaching and learning approaches, structuring content, and determining the competencies required to fit current situations and conditions.

Badan Standar Nasional Pendidikan (BSNP) or the National Education Standards Agency states that the primary educational goal for the 21st century is to embody the goals of the nation—producing blissful and prosperous citizens that are respected and can compete with those from other countries. As such, BSNP hopes to produce qualified HR who are self-reliant and willing to and capable of turning the ideals of the country into reality (Mukminan, 2014 and Daryanto and Karim, 2017). The Ministry of Education and Culture (MoEC) kept this goal in mind when it developed curricula for elementary, junior- and senior-high, and vocational schools. It also strives to incorporate 21st-century skills, scientific approaches, and learning and authentic assessment in curricula (Daryanto and Karim, 2017). Developing 21st-century skills leads to enhancing other growth-oriented learning skills, including to think, act, and live in the real world. According to Trilling and Fadel (2009), students must possess life and career, self-learning and innovation, and media information and technology skills in the 21st century.

A number of developed countries have been instigating reforms and innovations to make education more responsive to the need for 21st-century skills. And these reforms and innovations include promoting STEM education (Kuenzi, 2008). STEM education was first introduced by the U.S. National Science Foundation (NSF). Since then, it has been promoted in other countries. Indonesia has just recently introduced STEM education. In STEM education, science deals with the study of natural phenomena using observation and measurement to objectively explain environmental changes. Science, in this case, includes physics, biology, chemistry, and earth and space sciences. Technology, on the other hand, deals with human innovations that can be used to modify nature to help humans address their needs and improve their lives. Engineering, meanwhile, should address the need to acquire and apply science, social, economic, and practical knowledge to design and construct machines, tools, systems, materials, and processes that will benefit human beings. Finally, mathematics should provide people with knowledge of patterns and relationships and a means to understand the language of technology, science, and engineering (Firman, 2017).

Science in STEM education usually refers to natural sciences, but it can also be extended to include social sciences. STEM education actually aims to attract students to STEM-related careers that will characterise the 21st century. Focus on STEM education arises from the fact that learning mathematical and scientific concepts is no longer enough to become modern citizens who need to relate knowledge of these to technology and engineering (Chesky and Wolfmeyer, 2015). STEM education hopes to develop learners' problem-solving skills to deal with STEM-related challenges that can arise in their daily lives (Bybee, 2013).

STEM education will, however, become meaningless if the fields are separately reinforced. Hence, developing an educational approach that addresses all four components of STEM education to teach citizens to address real-life problems is necessary. For primary- and secondary-level students, this means improving their STEM literacy (Bybee, 2013). They should:

- Acquire the necessary knowledge, attitudes, and skills to identify problems, explain natural phenomena, and draw conclusions based on objective evidence

- Understand how imbibing the STEM discipline can enhance human knowledge and investigation and design skills

- Enhance awareness of how STEM can enhance the material, intellectual, and cultural environments

- Increase their willingness to participate in solving related issues such as energy conservation, environmental quality, and natural resource constraints as constructive, caring, and reflective citizens who rely on STEM-based ideas

STEM education requires teachers not only to equip students with scientific knowledge, but also develop their soft and critical-thinking skills or ability to analyse, assess, and improve their thinking (Paul, 2008). Following the STEM education approach can respond to a wide range of daily problems and challenges, including those related to population, economic, social, and political issues (Pattanapichet and Wichadee, 2015 and Forawi, 2016). In addition to addressing workforce issues, STEM education will also benefit the general public with a collective of citizens that have the skills and knowledge to live better lives by making the right personal decisions (Urban and Falvo, 2016).

Several studies on using technology to improve STEM education have been published. More in-depth research, however, is still needed to deepen our understanding of STEM education and enrich teaching and learning approaches to benefit the future workforce. This includes finding out if today's students have the four abilities mentioned earlier. The ability to solve STEM problems reflects students' critical-thinking skills because the act requires them to analyse and assess factors to resolve issues. Teachers can use the study's results to design a learning process that integrates all of the necessary STEM components.

This article features our preliminary research on STEM education in high school. We asked the students from West Java STEM questions to test how prepared they are for 21st-century careers. We analysed the results using the Rasch model via the Winstep software (Khotimah and Sri, 2014). Compared with other validation methods such as the classical test theory, the Rasch model can predict missing data based on a systematic response pattern. This will produce a more accurate statistical analysis of test results despite missing data.

Research Methodology

Some 301 11th-grade students from five elementary schools in West Java participated in the study. Students were asked to participate in a STEM learning activity then answer an online questionnaire prepared by teachers who underwent SEAQIS training on STEM education. The questions tested students' understanding, applying, analysing, and evaluating cognitive skills.

In this study, a Rasch measurement model software—Winstep version 3.73—was used for dichotomous responses (items with only two potential responses—true or false). The Rasch model transforms raw item difficulties and raw person scores to equal interval measures of logits on a line in a "metre stick." This helps determine how much attention should be given to each item, depending on its position relative to those of others. Chan, et al. (2014) cited seven advantages of using the Rasch model, including:

- It can evaluate if the items used are appropriate and if biases can be identified.
- Item calibration is not affected by a person's ability.
- The calibration standard error can be exploited to examine the accuracy of each item.
- The model can predict how difficult an item is and make it less so using a standard scale.
- Respondents' abilities can be compared with one another even if they do not answer the same items by using a common scale for ability estimation.
- The Chi-square of person fit can be used to assess the quality of measurement.
- The model can improve the construction and design of a test so assessment can be easily adjusted.

In the study, an instrument was deemed good if it had high validity and reliability scores. In the Rasch model, the indicators that should be observed to determine reliability are referred to as the Cronbach Alpha, which measure a person's and an item's reliability (Saad, et al., 2011). Table 11 shows the ranges of reliability measurements in the model. The indicators that should be observed to determine validity, meanwhile, are values of point measure correlation and MNSQ values (Mohamad, et al., 2015). The reliability scale used in the study was based on a rating scale instrument developed by Fisher (2007).

Table 11: Person and Item Measurement Reliability	
Reliability Measurement	**Range**
Poor	< 0.67
Fair	0.67–0.80
Good	0.81–0.90
Very Good	0.91–0.94
Excellent	> 0.94

Research Results and Findings

All of the data collected was analysed using the Rasch measurement model software, Winsteps version 3.37. Statistics was used to measure the test's reliability by analysing the consistency between items. A higher value indicates a strong relationship between items in the test whereas a small value indicates a weak relationship.

```
INPUT: 301 PERSON  10 ITEM  REPORTED: 301 PERSON  10 ITEM  2 CATS  WINSTEPS 3.73
--------------------------------------------------------------------------------

     SUMMARY OF 301 MEASURED PERSON
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT         OUTFIT   |
|          SCORE     COUNT    MEASURE   ERROR     MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN       2.7      10.0      -1.27     .89                                 |
| S.D.       1.7        .0       1.10     .34                                 |
| MAX.       9.0      10.0       2.32    1.86                                 |
| MIN.        .0      10.0      -3.62     .66      .69   -1.7    .44   -1.7 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .97 TRUE SD    .52  SEPARATION   .54  PERSON RELIABILITY  .22 |
|MODEL RMSE    .95 TRUE SD    .55  SEPARATION   .58  PERSON RELIABILITY  .25 |
| S.E. OF PERSON MEAN = .06                                                   |
-------------------------------------------------------------------------------
PERSON RAW SCORE-TO-MEASURE CORRELATION = .97
CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .37

     SUMMARY OF 10 MEASURED (NON-EXTREME) ITEM
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT         OUTFIT   |
|          SCORE     COUNT    MEASURE   ERROR     MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      81.1     301.0        .00     .15     1.00     .0   1.01     .2 |
| S.D.      27.6        .0        .56     .01      .10    2.0    .16    2.3 |
| MAX.     137.0     301.0        .77     .17     1.22    4.8   1.39    6.3 |
| MIN.      46.0     301.0      -1.05     .13      .84   -3.3    .80   -2.7 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .15 TRUE SD    .54  SEPARATION  3.60  ITEM   RELIABILITY  .93 |
|MODEL RMSE    .15 TRUE SD    .54  SEPARATION  3.66  ITEM   RELIABILITY  .93 |
| S.E. OF ITEM MEAN = .19                                                     |
-------------------------------------------------------------------------------
```

Figure 14: Output data from the instrument's analysis

Figure 14 shows the output data from the analysis of the instrument used in the study. The person's reliability score was 0.22 while the item's was 0.93. The person separation index was 0.54 while the item separation index was 3.60. Using the rating scale in Table 11, the person's reliability was poor. And based on Fisher's scale (2007), the person separation index value was also poor. The results reflect the instrument's low sensitivity to distinguish between high and low performers. This could be due to its low quality or the small number of test items (Rashidi, et al., 2014 and Abell and DeBoer, 2015). To increase a person's reliability, the students' ability ranges need to increase and more items should also be added to the test (Chen, et al., 2014).

Unlike the person reliability score, the item reliability and item separation index scores were good. This shows that the instrument worked well in terms of performing further analysis. The instrument can, however, categorise the items based on difficulty—lower, moderate, and upper—to become more reliable with regard to measuring given constructs.

The test items in the study were considered a reliable measuring instrument based on initial analysis. They were validated by experts from a university to see if they match the scientific concepts the study hopes to use. They were also analysed for accuracy. An item was considered valid if it corresponded to a scientific concept taught in class or is compatible with one in the science curriculum (Firman, 2000). Validation was a vital component of the study because it ensured the test's ability to evaluate students' ability to solve STEM problems. The validity test assessed the instrument in relation to set criteria, content, and construct (Swerdlik, 2009). The instrument should match the subject's content. Experts from a college were asked to evaluate the content validity of the questions. Construct validity allowed the users to determine if the scores of the instrument were significant, meaningful, useful, and purposive. Three misfit patterns were considered in assessing the construct validity of an item using point measure correlation and MNSQ.

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
10	137	301	-1.05	.13	1.22	4.8	1.39	6.3	A	.27	.45	58.1	64.9	S10
6	75	301	.07	.14	1.06	.9	1.10	1.0	B	.32	.37	73.5	74.3	S6
2	51	301	.64	.16	1.01	.1	1.07	.6	C	.31	.32	81.6	81.9	S2
5	52	301	.61	.16	1.05	.5	1.03	.3	D	.29	.32	82.0	81.6	S5
8	46	301	.77	.17	1.03	.3	1.02	.2	E	.29	.31	82.0	83.5	S8
1	96	301	-.33	.14	.98	-.5	.98	-.3	e	.42	.40	71.3	69.0	S1
9	91	301	-.24	.14	.97	-.4	.95	-.6	d	.42	.39	73.2	70.2	S9
7	103	301	-.46	.13	.95	-1.1	.93	-1.1	c	.46	.41	71.0	67.4	S7
3	62	301	.36	.15	.88	-1.5	.80	-1.7	b	.44	.34	82.0	78.3	S3
4	98	301	-.37	.14	.84	-3.3	.82	-2.7	a	.53	.40	77.2	68.5	S4
MEAN	81.1	301.0	.00	.15	1.00	.0	1.01	.2				75.2	74.0	
S.D.	27.6	.0	.56	.01	.10	2.0	.16	2.3				7.2	6.5	

Figure 15: Item-specific statistics

Point measure correlation statistics was then used to check if all of the items in the test worked towards achieving the same goal. Figure 15 shows that all of the test items had positive values, which indicated their acceptability (Bond and Fox, 2007). These indicated the measured items' parallelism with the test's construction. The MNSQ scores of all items were between 0.82 and 1.39. None of them needed to be modified or removed.

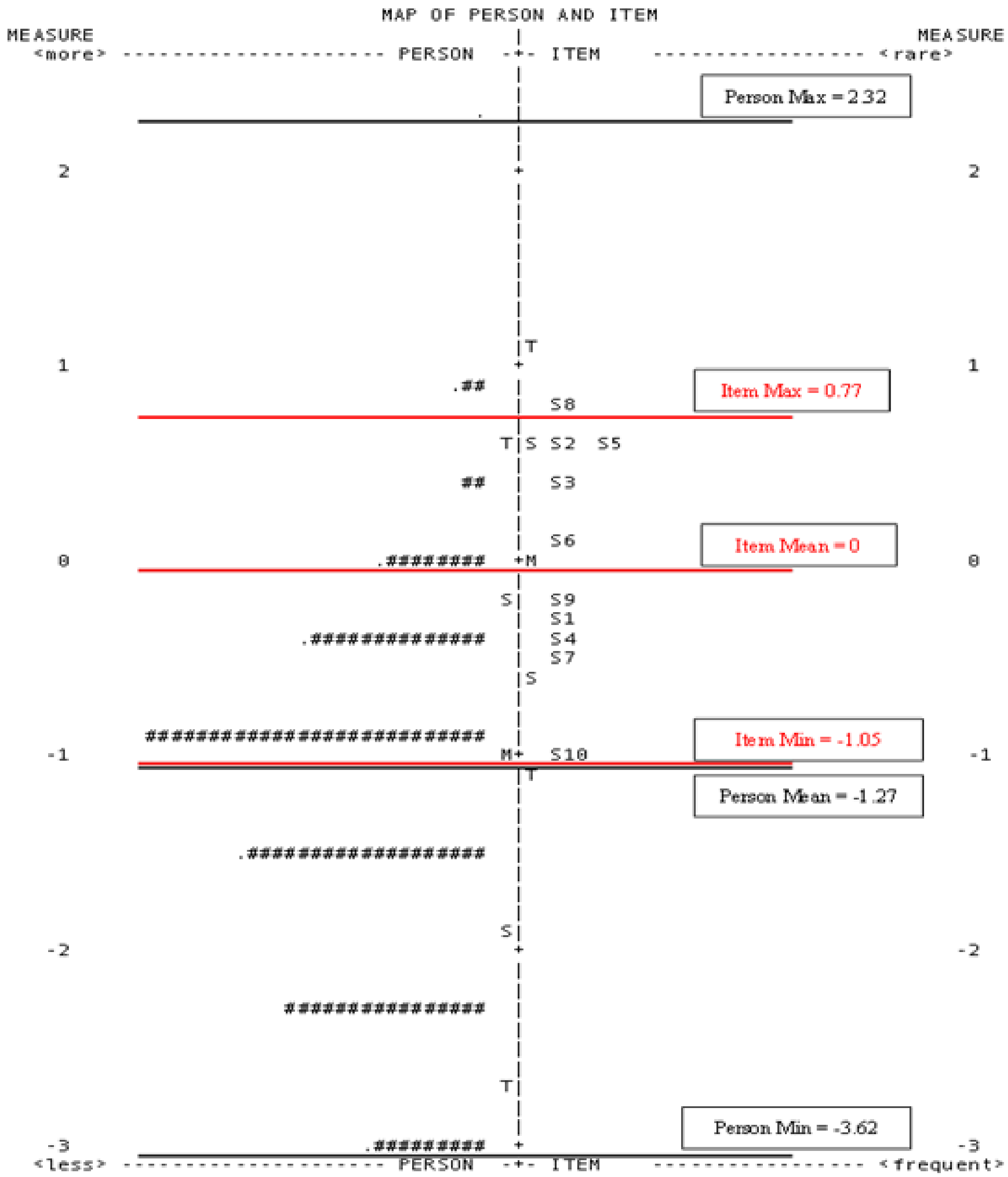

Figure 16: Variable map

In Figure 16, the number signs and periods on the left represented the students. Each number sign represented three students while each period stood for 1–2 students. The variable map, also known as the "Wright Map," correlates students' ability (person) with the level of difficulty of each given problem (item). Students were scored based on their ability. The less able they were, the lower they were on the scale. The items on the right were also arranged from easiest (lowest portion) to hardest (topmost portion). The more difficult a problem (item) was, the higher it was on the scale.

Zero referred to average student ability (left) and test difficulty (right). As shown, the average for student ability was lower than that for test difficulty, indicating that the tests were, on average, relatively difficult for the students. Most of the students (86.71%), were in fact, below average. The logit values that their ability to solve STEM problems was low because they were not able to correctly answer the questions.

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%
242	9	10	2.32	1.07	1.20	.5	3.00	1.7	-.62	.15	90.0	90.0
78	7	10	.91	.71	.89	-.3	.84	-.3	.45	.24	70.0	70.0
80	7	10	.91	.71	.89	-.3	.84	-.3	.45	.24	70.0	70.0
248	7	10	.91	.71	1.04	.2	1.28	.8	.08	.24	70.0	70.0
268	7	10	.91	.71	1.13	.5	1.37	1.0	-.08	.24	70.0	70.0
274	7	10	.91	.71	1.00	.1	1.24	.7	.14	.24	70.0	70.0
277	7	10	.91	.71	1.13	.5	1.37	1.0	-.08	.24	70.0	70.0
279	7	10	.91	.71	1.00	.1	1.24	.7	.14	.24	70.0	70.0
297	7	10	.91	.71	1.17	.6	1.40	1.1	-.13	.24	70.0	70.0
77	6	10	.44	.67	1.10	.6	1.21	.9	.03	.26	70.0	63.5
144	6	10	.44	.67	1.03	.2	1.01	.1	.21	.26	50.0	63.5
							shortened					
229	1	10	-2.32	1.07	1.00	.3	.80	.1	.22	.17	90.0	90.0
230	1	10	-2.32	1.07	.98	.3	.74	.1	.27	.17	90.0	90.0
232	1	10	-2.32	1.07	1.17	.5	1.98	1.1	-.36	.17	90.0	90.0
233	1	10	-2.32	1.07	1.17	.5	2.03	1.1	-.38	.17	90.0	90.0
240	1	10	-2.32	1.07	1.17	.5	2.03	1.1	-.38	.17	90.0	90.0
243	1	10	-2.32	1.07	1.03	.3	.90	.2	.14	.17	90.0	90.0
260	1	10	-2.32	1.07	1.03	.3	.90	.2	.14	.17	90.0	90.0
290	1	10	-2.32	1.07	1.17	.5	2.03	1.1	-.38	.17	90.0	90.0
7	0	10	-3.62	1.86	MINIMUM MEASURE				.00	.00	100.0	100.0
11	0	10	-3.62	1.86	MINIMUM MEASURE				.00	.00	100.0	100.0
							shortened					
194	0	10	-3.62	1.86	MINIMUM MEASURE				.00	.00	100.0	100.0
205	0	10	-3.62	1.86	MINIMUM MEASURE				.00	.00	100.0	100.0
210	0	10	-3.62	1.86	MINIMUM MEASURE				.00	.00	100.0	100.0
223	0	10	-3.62	1.86	MINIMUM MEASURE				.00	.00	100.0	100.0
239	0	10	-3.62	1.86	MINIMUM MEASURE				.00	.00	100.0	100.0
245	0	10	-3.62	1.86	MINIMUM MEASURE				.00	.00	100.0	100.0
270	0	10	-3.62	1.86	MINIMUM MEASURE				.00	.00	100.0	100.0
271	0	10	-3.62	1.86	MINIMUM MEASURE				.00	.00	100.0	100.0
295	0	10	-3.62	1.86	MINIMUM MEASURE				.00	.00	100.0	100.0
MEAN	2.7	10.0	-1.27	.89	.99	.1	1.01	.1			75.2	74.0
S.D.	1.7	.0	1.10	.34	.16	.6	.36	.7			12.7	9.6

Figure 17: Measure order

The variable map showed that the students had the greatest difficulty in solving item S8. This question was related to dengue fever. It provided experimental data on how the pH level of water is related to mosquito volume. The students' failure to incorrectly answer the question indicates their lack of knowledge on or even interest in solving the dengue problem. The students, meanwhile, found item S10 easy to answer. They were asked to interpret a chart and make a prediction based on the information it showed. This indicated that most students can interpret data shown in a chart. Items S2 and S5 had the same level of difficulty. One of them need to be eliminated or improved to better gauge the students' ability.

The students were ranked according to their ability, from the most to least able. Based on the data, only one student (+2.32) had a high STEM problem-solving ability while 29 students did not. The STEM problems used were contextual in nature, which means they affected the students' daily lives. The test results showed that the students still lacked the ability to apply what they learn in real life. They lacked HOTS.

Conclusion

The study revealed that the students still lacked the ability to solve STEM problems. The Rasch model was used to analyse the students' answers to 10 multiple-choice questions related to STEM education. The resulting variable map showed that most of the students had very poor STEM problem-solving ability. This could be due to their unfamiliarity with STEM-related concepts and so teachers would do well to educate them on such.

References

1. Bambang Sumintono and Wahyu Widhiarso. (November 2014). "Aplikasi Model Rasch Untuk Penelitian Ilmu-Ilmu Sosial (Edisi Revisi)." Retrieved from https://www.researchgate.net/publication/268688933_Aplikasi_Model_Rasch_untuk_Penelitian_Ilmu-Ilmu_Sosial_edisi_revisi.

2. Bambang Sumintono and Wahyu Widhiarso. (September 2015). "Aplikasi Pemodelan Rasch pada Assessment Pendidikan." Retrieved from https://www.researchgate.net/publication/282673464_Aplikasi_Pemodelan_Rasch_pada_Assessment_Pendidikan.

3. Benjamin Wright and Mark Stone. (1999). "Measurement Essentials (Second Edition)." Retrieved from https://www.rasch.org/measess/me-all.pdf.

4. Cari F. Herrmann-Abell and George E. DeBoer. (April 2015). "Using Rasch Modelling to Explore Students' Understanding of Elementary-School Ideas About Energy." Retrieved from https://files.eric.ed.gov/fulltext/ED560007.pdf.

5. Daryanto and M. T. Syaiful Karim. (2017). "Pembelajaran Abad 21." Gava Media.

6. Harry Firman. (2000). "Penelitian Hasil Belajar Dalam Pengajaran Kimia." Jurusan Pendidikan Kimia FPMIPA UPI.

7. Harry Firman. (2017). "Pendidikan STEM Untuk Pembelajaran Kimia Abad Ke 21." Paper presented during the "Training Course on Subject Knowledge Enrichment in Chemistry" at SEAQIS, Bandung, Indonesia.

8. Jeffrey J. Kuenzi. (21 March 2008). "STEM Education: Background, Federal Policy, and Legislative Action." Retrieved from https://fas.org/sgp/crs/misc/RL33434.pdf.

9. Khotimah Marjiastuti dan Sri Wahyuni. (2014). "Analisis Kemampuan Peserta Didik Dengan Model Rasch." Retrieved from https://conf.unnes.ac.id/index.php/snep/II/paper/viewFile/175/67.

10. Mark Wolfmeyer and Nataly Z. Chesky. (2015). "Philosophy of STEM Education: A Critical Investigation." Palgrave Macmillan.

11. Michael J. Urban and David A. Falvo. (2016). "Improving K–12 STEM Education Outcomes through Technological Integration." Retrieved from https://eric.ed.gov/?redir=http%3a%2f%2fdx.doi.org%2f10.4018%2f978-1-4666-9616-7.

12. Mimi Mohaffyza Mohamad Nor, Lisa Sulaiman Lai, Chee Sern Kahirol, and Mohd Salleh. (24 August 2015). "Measuring the Validity and Reliability of Research Instruments." Retrieved from https://www.sciencedirect.com/science/article/pii/S1877042815047771.

13. Mohd Nor Rashidi, Rawshan Ara Begum, Mazlin Bin Mokhtar, and Joy Jacqueline Pereira. (2014). "The Applications of Rasch Measurement Model for Calibrating a Survey Instrument to Analyse Criteria for Sustainable Construction." Retrieved from http://dx.doi.org/10.12944/CWE.9.3.14.

14. Mukminan. (24 April 2014). "Strategi Menyiasati Pendidikan Abad 21." Retrieved from http://staffnew.uny.ac.id/upload/130682770/penelitian/ba-22semnas-upiupistr-menyiasati-pddk-abad-xxi.pdf.

15. Richard Paul and Linda Elder. (2006). "The Miniature Guide to Critical Thinking: Concepts and Tools." Retrieved from https://www.criticalthinking.org/files/Concepts_Tools.pdf.

16. Rodger W. Bybee. (2013). "The Case for STEM Education: Challenges and Opportunities." NSTA Press.

17. Rohaizah Saad, Rushami Zein Yusuff, Zakaria Abas, Azrilah Abd Aziz, and Mohd Saidfudin Masodi. (2011). "Validating the International Organization for Standardization (ISO) 9000 Construct of Measurement Instrument through the Application of the Rasch Model." Retrieved from http://journal.sbm.itb.ac.id/index.php/ajtm/article/download/160/147.

18. Ronald Jay Cohen and Mark E. Swerdlik. (2009). "Psychological Testing and Assessment (Seventh Edition)." McGraw-Hill.

19. Shiau Wei Chana, Zaleha Ismail, and Bambang Sumintono. (September 2013). "A Rasch Model Analysis on Secondary Students' Statistical Reasoning Ability in Descriptive Statistics." Retrieved from https://umexpert.um.edu.my/file/publication/00013268_112483.pdf.

20. Trevor G. Bond and Christine M. Fox. (2007). "Applying the Rasch Model: Fundamental Measurement in Human Sciences (Second Edition)." Retrieved from https://researchonline.jcu.edu.au/9907/2/9907_Bond_%26_Fox_front_pages.pdf.

21. William P. Fisher, Jr. (2007). "Rasch Measurement: Transactions of the Rasch Measurement SIG American Educational Research Association." Retrieved from https://www.rasch.org/rmt/rmt211.pdf.

22. Fasawang Pattanapichet and Saovapa Wichadee. (October 2015). "Using Space in Social Media to Promote Undergraduate Students' Critical-Thinking Skills." Retrieved from http://dergipark.gov.tr/download/article-file/156696.