

---

## **Validation of Science, Technology, Engineering, and Matemathic Based Diagnostik Assessment on Natural Resource Material for Phase B Elementary School Students using Rasch Model**

**Jihan Nadhifa Dewi, Siti Patonah\*, Sukamto**

Universitas PGRI Semarang, Semarang, Indonesia

\*Email: [sitifatonah@upgris.ac.id](mailto:sitifatonah@upgris.ac.id)

---

### **Article History:**

Received date: March 10, 2023  
Received in revised from: June 25, 2023  
Accepted date: July 21, 2023  
Available online: July 29, 2023

### **Citation:**

Dewi, J.N., Patonah, S., & Sukamto. 2023. Validation of science, technology, engineering, and matemathic based diagnostik assessment on natural resource material for phase B Elementary School students using rasch model. *Jurnal Pendidikan Sains Indonesia (Indonesian Journal of Science Education)*, 11(3):654-667.

### **Abstract.**

One of the ways to design learning the implementation of the independent curriculum is assessment diagnostic. This study aims to develop STEM-based questions to measure students' critical thinking skills. The questions are used for the initial assessment of science subject material "Natural Resources" phase B elementary school. The type of research used is research and development (R&D) with the ADDIE development model, consisting of the analysis, design, developmen, implementation, and evaluation stages. The school students in the city of Semarang with a total of 258 students consisting of 126 male students and 132 female students, using convenience sampling. The items' validation and reliability analysis techniques used the Aiken index and the RASCH model. The results showed that all items had a high validity category based on the following aspects: Outfit items, measure, test difficulty level, person and item reliability. There are several phase bias items, where students in phase C benefit more than students in phase B. The average respon ability is above the average item difficulty level. The conclusion is that the items developed can be used as initial assessments in science subjects.

**Keywords:** Assessment Diagnostik, STEM, RASCH Model, Elementary School

---

## **Introduction**

Learning the new paradigm ensures that the practice of learning is learner-centered (Azis et al., 2022) The new learning paradigm provides educators the flexibility to formulate learning designs and assessments according to the characteristics and needs of students (Phil, 2021). In learning the new paradigm, the learning and assessment cycles are interrelated so that students can achieve the expected competencies. Using appropriate assessments for teachers will determine the weaknesses in learning so that teachers can improve methods in the learning process (Rosnaeni, 2021).

Assessment can be interpreted as comparing or applying measurement results to provide value to the object (Zainal, 2020). Students can carry three aspects of the realm of assessment: knowledge, attitudes, and skills (Jeanne & Prendergast, 2020) Applying diagnostic assessments before designing learning is a hallmark of an independent curriculum. The main direction of learning in the independent curriculum is that students can take advantage of and hone their competencies in their social life (Nurmasyitah et al., 2023). The independent learning curriculum is specifically designed to give the right to

learn independently (Inayati, 2022). The independent curriculum is expected to have diverse extracurriculars (Purnawanto, 2022). An independent curriculum makes learning more optimal, giving students enough time to understand concepts and strengthen their competencies.

Based on observations from several elementary schools in the city of Semarang and also as a first step in implementing an independent curriculum, researchers develop an assessment of questions. The questions developed are based STEM. The experience students gain in STEM learning is that students can gain knowledge in science, technology, engineering, and mathematics. That knowledge can then be used to solve real-life problems and use them meaningfully in life (Jang et al., 2022). The benefits of implementing STEM education are that it can improve critical and creative thinking skills. In addition, students can be more logically creative, innovative, productive, and directly related to real conditions (Wang & Wang, 2023).

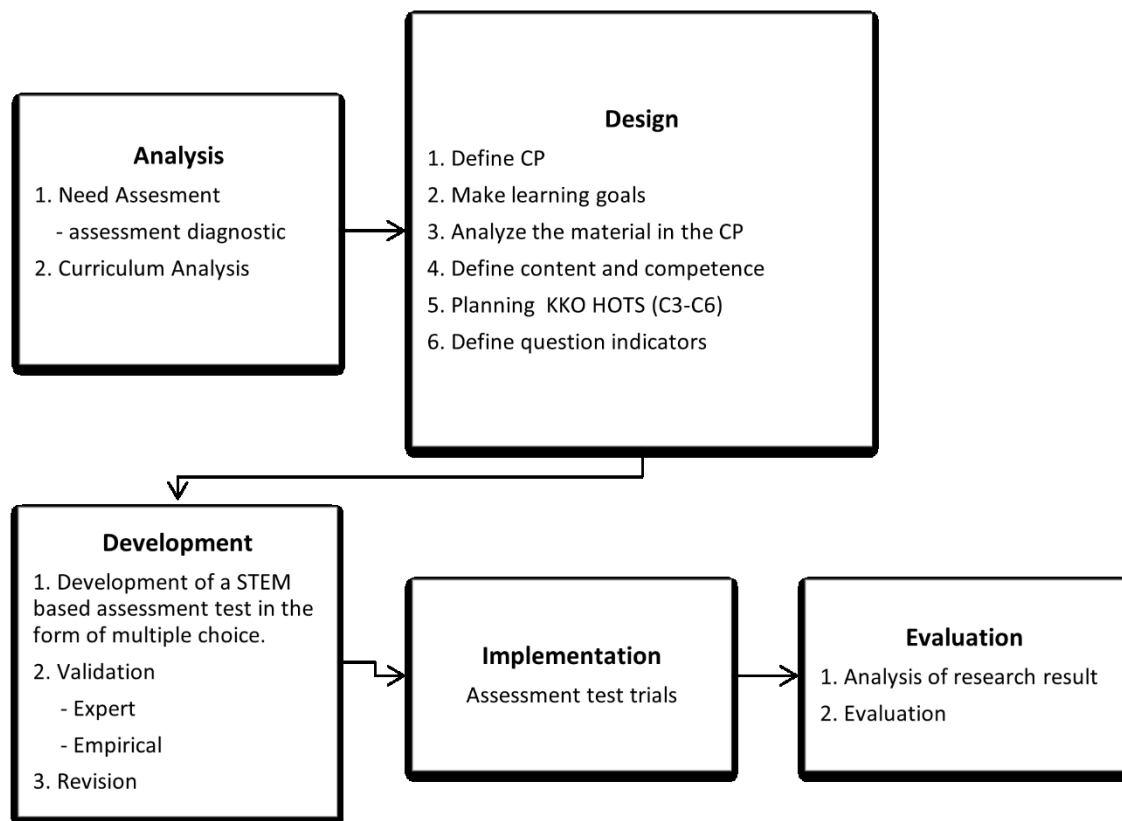
Science is commonly referred to as natural science. Science studies how this universe works, so science can be called a science that studies nature itself. In the 21st century, four skills must be met, namely 1) digital era literacy, 2) inventive thinking, 3) effective communication, and 4) high productivity (Patonah et al., 2020). Scientific literacy is part of literacy which include knowledge and understanding of scientific concepts and processes (Cai-Ting et al., 2020). Science skills are a must-have in the 21st century and can be realized in everyday life. The challenge in learning in the 21st century is integrating pedagogy, technology, and assessment and involving teachers, students, and leaders as stakeholders (Sustiningsih et al., 2021).

Previous research has proven that developing STEM-based test assessment instruments is feasible for measuring students' creative thinking ability (Agustina, 2022). STEM is a modern approach that integrates all of its aspects to solve problems that arise 21st-century life (Hebebcı & Usta, 2022). STEM learning aims to solve problems in everyday life by applying it in schools whose learning subjects combine the knowledge, skills, and attitudes students possess (Subayani, 2022).

Applying the right STEM approach model can affect students' success in understanding the material. It can motivate students to find learning concepts (Izzati et al., 2019). Thus, developing this instrument is necessary as an initial assessment of implementing an independent curriculum. This study aims to develop STEM-based questions on natural resource materials for phase B elementary school students as the first step in implementing an independent curriculum.

## **Methods**

This form of research is research and development (R&D). Research and development is a research method used to develop or validate products for teaching and learning (Nugroho & Airlan, 2020). This research was used to develop an assessment test using questions to measure the critical thinking skills of phase B elementary school students on natural resource materials. This research refers to the model developed by Dick and Carry (1996), analysis, design, developmen, implementation, and evaluation (ADDIE) models. The subjects involved in this study were grade III and IV students at SD N in the Semarang city area, with a total of 258 students consisting of 126 male and 132 female students. ADDIE's research and development procedures in detail are shown in flow chart.



The first stage is the analysis stage. Before a diagnostic assessment is developed, a needs analysis and material analysis are first carried out. The analysis was carried out on primary school teachers and elementary school students. At this stage, its activities include collecting information such as teacher needs analysis, student needs analysis, cognitive analysis, and analysis of critical thinking skills. Some of the main things obtained at this stage are 1) implementing STEM-based *diagnostic assessments* has not been applied to phase B natural resource materials, 2) most questions measure aspects of memory that cannot be used to measure students' critical thinking ability.

The second stage is the design or role, carried out after analyzing the existing problem. At this stage, researchers begin to design products to be developed to measure product quality. The activities carried out include 1) determining learning outcomes (LO) based on phases determined based on independent curriculum documents, 2) determining material on learning that has been determined in LO, 3) recognizing competencies and content contained in LO, 4) planning the use of active verbs that are comparable to HOTS questions, starting from C4 to C6 at a predetermined taxonomic level, 5) determine the indicators of the questions, 6) make the question items.

The third stage is the development of the problem. The development product in the form of question items is then validated by experts and practitioners before being tested on students. The product validation results of the question items are revised by expert advice and input. The product of the question item can be implemented if the product has been declared. Validation is carried out through 2 things, namely expert and empirical

validation. Expert validation is carried out to review the accuracy of question items that can be seen in terms of material (substance), construction, and language. Seven validator experts, including lecturers and principal practitioners, carry out expert validation. The scores given by the expert use the criteria shown in **Table 1**.

**Table 1.** Validation Criteria

Grade	Description
5	Highly relevant
4	Relevant
3	Quite Relevant
2	Less Relevant
1	Irrelevant

Analysis data from expert validation was calculated using the Aiken index obtained from seven predetermined validators. The Aiken validity index can be calculated using the formula:

$$s = r-1 \quad (1)$$

$$V = \frac{\sum s}{n(c-1)} \quad (2)$$

Remark:

V= rater agreement index regarding validation of question items

r= rater assessment

s= difference of rater reduced by the minimum value of the rater

n= number of appraisers

c= maximum value of validity

Seven experts further validate the questions compiled with the criteria indicated in Table 2.

**Table 2.** Aiken Index Criteria

Score	Information
>0.4	Low
0.4-0.8	Medium
<0.8	High

Furthermore, the question item is empirically validated. Empirical validation is carried out to test the accuracy of question items and the reliability of instruments based on trials of samples determined in the study (Setemen, 2018). The question item validation test was carried out using the RASCH model with the help of the Winstep 3.73 application. The criteria used to check the suitability of the question item so that it can be said to be appropriate or inappropriate are by looking at the values of outfit mean square (MNSQ), Outfit Z-standard (ZSTD) and point measure correlation (Pt Mean Core). The criteria for items said to be valid can be seen in Table 3 (Arnold et al., 2018; Ramadhani & Fitri, 2020; Sumintono & Widhiarso, 2015).

**Table 3.** Question Item Validation Criteria

Criteria	Koefisien	Value
Outfit MNSQ	$0,5 < \text{MNSQ} < 1,5$	Accepted/Valid
Outfit ZSTD	$-2,0 < \text{ZSTD} < +2.0$	Accepted/Valid
Pt. Mean Core	$0,4 < \text{Pt. Mean Core} < 0,85$	Accepted/Valid

The fourth stage is the implementation stage: implementing a diagnostic assessment to be tested on elementary school students. The student in Phase B in the independent curriculum (source) in 2 elementary schools in Semarang City using convenient sampling techniques. Item questions were tested in 2 elementary schools in Semarang city in grades 3 and 4. A STEM-based assessment test on natural resource materials was tested on 258 students with different abilities. Trials were conducted to determine the extent to which the level of question items can achieve the desired goal. The quality of the question items developed will be tested empirically through trials.

The last stage is the evaluation stage. In this stage, the question items that have been tested are then analyzed using the Winstep 3.73 application. Each question item is analyzed based on reliability, validity, difficulty level, and bias on certain variables.

## Results and Discussion

The *assessment test* instrument developed is an elaboration of each predetermined learning outcome. The elaboration of learning outcomes produces learning objectives to be used as a reference for making questions. Problem indicators are based on content (material) aspects, competency aspects, and taxonomic levels. Each indicator created is then divided into two questions with different taxonomic levels. The taxonomic level for STEM is C4 to C6 because it is indicated by a question with a high level of intelligence (Damanik & Irfandi, 2021; Khaldun et al., 2020; Kristanto & Setiawan, 2020; Ndiung & Jediut, 2020; Widarta & Artika, 2021; Yuliandini et al., 2019). Any question containing STEM can measure students' critical thinking ability.

Based on predetermined learning outcomes produces three learning objectives. The learning objectives were developed into 7 question indicators. Each question indicator compiled two questions according to the level of taxonomy bloom Table 4.

**Table 4.** The learning outcomes elaboration matrix becomes a problem indicator

Learning Outcomes	Objective Learning	Question Indicator	Order of Question	Key Answer
Learners can identify problems related to the preservation of natural resources in the surrounding environment	Students can conclude the utilization of natural resources in the surrounding environment.	Presented in a case, students can conclude human activities as an effort to preserve living things	1	C
			2	A
		Presented statements, learners can conclude the role of alternative energy in Preservation Efforts	11	B
			12	A

Learning Outcomes	Objective Learning	Question Indicator	Order of Question	Key Answer	
and their relation to efforts to conserve living things		Living things			
		Presented with questions, students can conclude the benefits of alternative energy for the preservation of living things	13	B	
	Students can relate efforts to preserve living things in the surrounding environment	Presented with questions, students can conclude human activities as an effort to preserve living things	14	A	
			3	D	
		Presented a picture, students can conclude the activities of preserving living beings	4	C	
			7	B	
		Presented pictures, students can deduce activities that undermine the sustainability Living things	8	B	
			9	A	
		Learners can diagnose the reasons for efforts to preserve living things.	Presented some efforts to preserve living beings, students can diagnose the reason for the effort that Done	10	C
				5	A
			6	A	

The results of the analysis of question items by expert experts calculated using the Aiken Index are shown in Table 5.

**Table 5.** Expert Validation Aiken Index

Question	Indeks Aiken		
	A	B	C
1	0.85	0.85	0.84
2	0.85	0.82	0.82
3	0.86	0.83	0.82
4	0.86	0.77	0.82
5	0.84	0.78	0.82
6	0.82	0.78	0.82
7	0.82	0.82	0.80
8	0.83	0.77	0.80
9	0.83	0.80	0.80
10	0.84	0.80	0.80
11	0.79	0.80	0.79
12	0.79	0.80	0.79
13	0.82	0.82	0.81
14	0.83	0.80	0.78

Based on the Aiken Index table validated by expert validators, column A is the result of validation based on the material (substance), column B is the result of validation based on construction, and column C is the result of validation based on language. The study results showed that out of 14 STEM-based questions on natural resource materials for phase B students, they showed good validation based on the validity criteria, the value

stated in the "valid" category ( $>0.4$ ) (Azhar et al., 2020). The values obtained are 0.77-0.86 regarding material, construction, and language. The validity test of items used to determine the cause of the question have low criteria (Wulandari et al., 2022).

After being validated, there are several suggestions for improvement from expert experts, namely the formulation of the subject matter that is not clear and clear, spelling improvements, and the use of appropriate EYD in the question item. Furthermore, improvements were made by formulating the subject matter clearly and clearly and using Indonesian properly and correctly according to the appropriate spelling and EYD.

In the next stage, 14 questions were developed and tested on elementary school students with 258 respondents, 126 men, and 132 women. It aims to determine the suitability of the model (question validity, difficulty level of the question, and reliability of the question). The questions developed were tested in 2 elementary schools in Semarang. Table 6 shows a summary of the data analysis results with the Winstep application's help.

**Table 6.** Summary Statistic

	Aspect	Value
Model Error	Person	0.87
	Item	0.15
Reliabilitas	Person Reliability	0.68
	Item Reliability	0.89
	Alpha Cronbach	0.81

In the RASCH model, the difference in power value can be seen from the error standard. Nilai model SE  $> 0.5$  identifies that the different power of the item is good. In the SE model, 0.5-1, the difference power is categorized as quite capable of distinguishing. In the SE model  $< 1$ , the difference power is inadequate or unable to distinguish (Purniasari et al., 2021). Based on the results of the research conducted, it shows that the SE items identify that the differential power of the items is good. Meanwhile, the SE person identified that the discriminating power generated was quite capable of discriminating.

Table 6 shows that the criteria of person reliability are sufficient, item reliability is good, and Cronbach's alpha value is good. The criteria are obtained from the view (Sumintono & Widhiarso, 2015) that in determining *item* reliability and *person reliability* based on the following values:  $<0.67$  (weak); 0.67-0.80 (sufficient); 0.81-0.90 (good); 0.91-0.94 (very good);  $>0.94$  (unique). This value shows that the quality of the question items shows excellent results with good reliability, while the quality of human resources is in the category of sufficient. In addition, the value of alpha Cronbach indicates that the interaction between a person and an item as a whole is a good value. Reliability test was carried out using the cronbach alpha formula with criteria of 0.6 (Tursinawati & Widodo, 2019). On the value of reliability or internal coefficient  $> 0.7$  indicates the questionnaire used is reliable (Puspita et al., 2021).

**Table 7.** Item Measure

Nomor Soal	Measure	Model S. E.	INFIT		OUTFIT		PT-Measure Corr
			MNSQ	ZSTD	MNSQ	ZSTD	
1	-0.06	0.15	0.90	-1.9	0.94	-0.8	0.55
2	0.35	0.15	0.98	-0.4	0.98	-0.2	0.57
3	0.45	0.15	1.09	1.8	1.11	1.3	0.54
4	-0.42	0.16	0.92	-1.4	0.89	-1.3	0.51
5	0.12	0.15	0.92	-1.5	0.89	-1.5	0.57
6	0.42	0.15	1.02	0.5	0.99	0.0	0.57
7	0.31	0.15	1.02	0.3	1.00	0.0	0.55
8	-0.82	0.16	1.12	1.7	1.17	1.6	0.38
9	0.69	0.16	0.94	-1.0	1.03	0.4	0.62
10	-0.56	0.16	1.02	0.4	1.11	1.2	0.45
11	-0.09	0.15	1.01	0.3	1.06	0.8	0.51
12	-0.46	0.16	0.89	-1.8	0.86	-1.7	0.52
13	0.66	0.16	1.06	1.1	1.14	1.5	0.57
14	-0.59	0.16	1.03	0.5	1.03	0.4	0.45

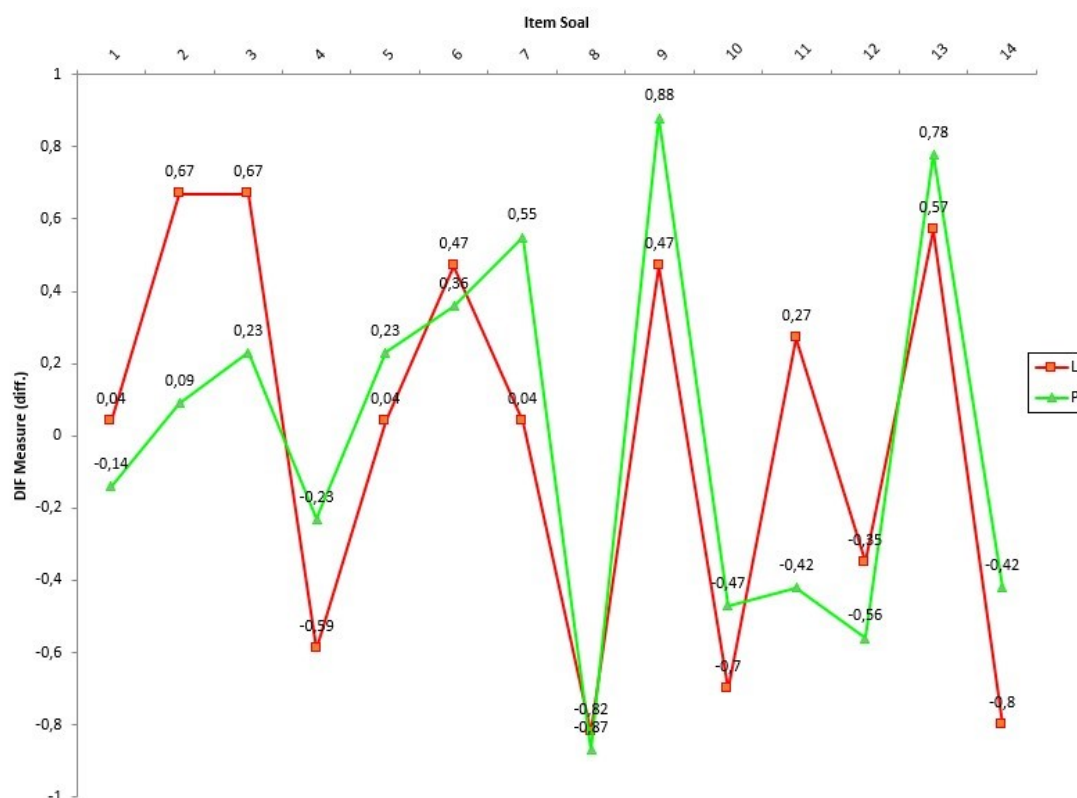
Table 7 summarizes the RASCH analysis of the Winstep 3.73 application-assisted model, data obtained from 258 respondents and 14 question items. Based on the measured value obtained, seven questions with the difficult item category, and seven with the easy item category. The difficulty level of the question items is reviewed according to (Sumintono & Widhiarso, 2015: 70), which is grouped into four categories, namely < -1 (specific items); -1 to 0 (easy items); 0 to 1 (difficult items); >1 (complicated items).

The analysis results of the problem instruments developed to show that they are valid. This is confused with the MNSQ infit value in the data analysis results, which shows  $1.12 \pm 0.89$ . the match conditions can be seen if the MNSQ INFIT value is 0.77-1.33 (Suparman, 2020). Thus it can be concluded that the STEM-based question instrument developed has been valid. The outfit mean squared (outfit MNSQ) value of 0.86 to 1.17 indicates the fit criteria. The match provision is between  $0.5 < \text{MNSQ} < 1.5$ . The test instruments are appropriate to measure students' ability to think critically about natural resource materials.

The distribution of the question item is considered a misfit or unfit if it meets one or both of the following conditions. The first condition is that the MNSQ outfit value value is between 0.5 and 1.5; ZSTD Outfit values are between -2.0 and 2.0; and the item correlation value with a total score point measure correlation located between 0.4 to 0.85 (Sumintono & Widhiarso, 2014). Testing the validity of question items using Pt. measure corr values, with valid criteria if they meet the range of  $0.4 < \text{Pt. Mean Corr} < 0.85$ . So it means there is a polarity item because the value obtained does not meet the required criteria range (Ramadhani & Fitri, 2020). In question item number 8 with the value of Pt. Mean Corr is 0.38. It is stated as an invalid question. While the MNSQ and ZSTD outfit values in Table 7. indicate values of 0.86-1.17 and -1.7-1.6 which means that the item of the developed question is valid or accepted.

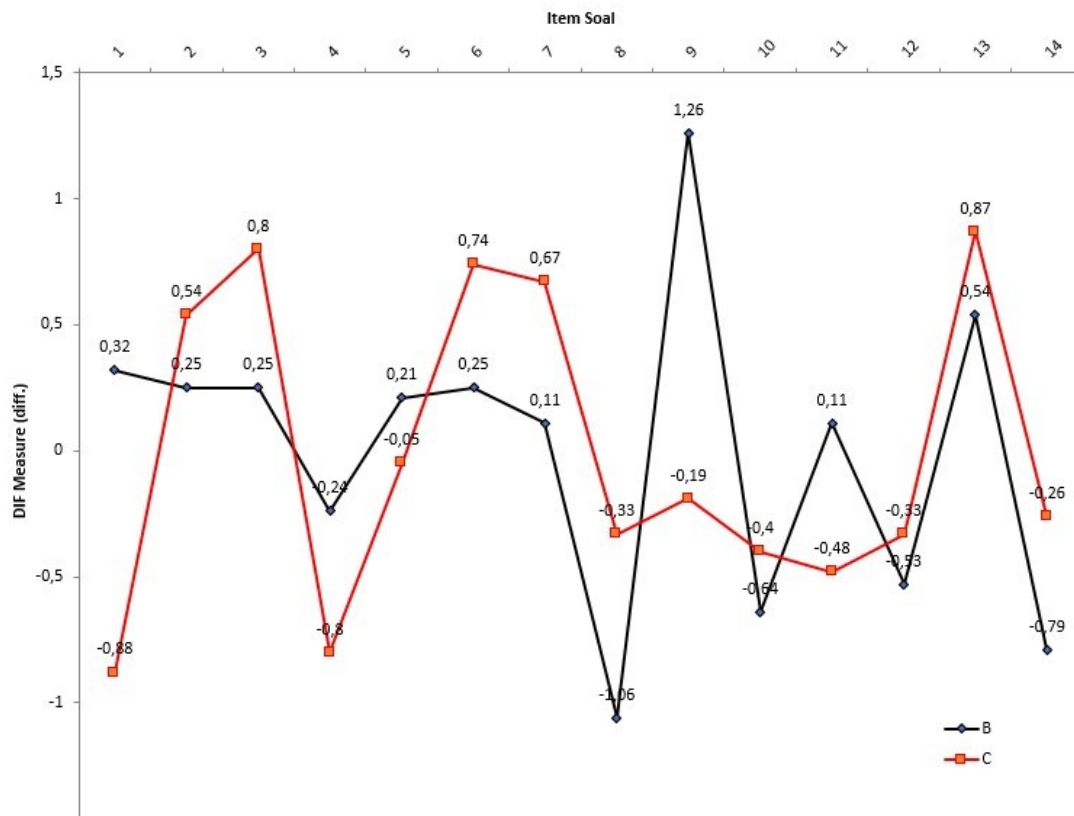


Items and measurement instruments that are biased if an item is more partial to one individual with specific characteristics and on the other side of an individual with an opposition character are harmed. The bias of a test can be interpreted as invalidity or sistematic error in measuring members of a group under study (Darmana et al., 2021). A question item is considered biased if the probability value is below 5% (0.05) (Burhani et al., 2022). Based on Figure 1 and Figure 2 shows the level of distribution in answering questions based on gender and phase.



**Figure 1.** DIF significant level graph by gender

Based on Figure 1, it can be seen the distribution of the difficulty level of the question by gender. The test takers consisted of two different groups, namely the male and female groups. Character identification needs to be done because it can affect the achievement of learning outcomes. The differential item functioning (DIF) value is used to find the instrument's bias. Figure 1 shows that male and female students have no bias and have an average ability not to answer questions well in doing the tested questions.



**Figure 2.** DIF significance level graph by phase

Based on Figure 2, it can be seen that the questions that experience bias (DIF) are questions number 1 and 9. The problem is infected with DIF because the probability value is below 0.05. it can be seen in the graph that the question item is easier for phase C students to do than for phase B students. Meanwhile, a curve close to the upper limit shows a question item with a great difficulty level (question number 9) (Aprilia et al., 2020).

## Conclusion

The research produced 14 stem-based questions on natural resource materials for phase B elementary school students with good quality. This development is used as an initial assessment in implementing an independent curriculum to measure students' critical thinking skills. Question items is carried out with two kinds of validation: 1) validation was carried out by seven validator experts with the acquisition of expert material, construction, and language using the Aiken index, 2) empirical validation performed by RASCH modeling using the help of Winstep 3.73 applications. Produce valid question items so that they can be used to measure students' critical thinking skills in phase B for science subjects. Further research can be done to determine the effectiveness of the questions in determining the map of students' abilities so that it can be used as a reference to carry out differentiated learning in elementary schools.

## Acknowledgement

The Researcher would like to thank the Ministry of Education, Culture, Research and Technology Directorate of Research, Technology and Community Service Directorate General of Higher Education, Research and Technology Contract Number: 158/E5/PG.02.00.PT/2022 dated 10 May 2022, 0026/LL6/PB/AK.04/2022 dated 11 May 2022.

## References

- Aprilia, N., Susilaningsih, E., Priatmoko, S., & Kasmui. 2020. Desain instrumen tes pemahaman konsep berbasis hot dengan analisis model rasch. *Chemistry in Education*, 9(2):1–8. <https://journal.unnes.ac.id/sju/index.php/chemined>.
- Arnold, J.C., Boone, W.J., Kremer, K., & Mayer, J. 2018. Assessment of competencies in scientific inquiry through the application of rasch measurement techniques. *Education Sciences*, 8(4): <https://doi.org/10.3390/educsci8040184>.
- Azhar, H., Fisika, P., Tarbiyah, F., Uin, K., & Makassar, A. 2020. Pengembangan media pembelajaran majalah mitigasi angin puting beliung terinterkoneksi ayat-ayat al-qur'an dalam pembelajaran fisika. *Jurnal Pendidikan Fisika*, 8(2):2355–5785. <http://journal.uin-alauddin.ac.id/index.php/PendidikanFisika>.
- Azis, F., Kaharuddin, K., Arifin, J., Yumriani, Y., Nawir, M., Nursalam, N., Quraisy, H., Rosa, I., Nuramal, N., & Karlina, Y. 2022. Pendampingan penguatan model pembelajaran paradigma baru bagi guru-guru sekolah muhammadiyah di kecamatan bontonompo selatan. *Jurnal Abdimas Indonesia*, 2(4):515–523. <https://doi.org/10.53769/jai.v2i4.337>.
- Burhani, S.N., Hakim, A., Hadisaputra, S., & Burhanudin. 2022. Analisis media pembelajaran phet simulations berbasis laboratorium virtual terhadap minat belajar kimia selama masa pandemi covid-19. *Chemistry in Education*, 5(2):194-201. <https://jurnalkip.unram.ac.id/index.php/CEP/article/view/3062/2727>.
- Cai-Ting, W., Liu, C.C., Chang, H.Y., Chang, C.-J., Chang, M.H., Chiang, S.H.F., Yang, C.W., & Hwang, F.K. 2020. Simulation and its relation to school science achievement and scientific literacy. *Komouter & Education*, 149. <https://doi.org/https://doi.org/10.1016/j.compedu.2020.10383>
- Damanik, F.C. & Irfandi. 2021. Pengembangan instrumen tes berbasis higher order thinking skill (HOTS) bentuk pilihan ganda berdasarkan taksonomi bloom revisi pada materi gelombang bunyi. *Jurnal Inovasi Pembelajaran Fisika*, 10(1):14–20. <http://jurnal.unimed.ac.id/2012/index.php/inpafi>.
- Darmana, A., Sutiani, A., Nasution, H.A., Ismanisa, I., & Nurhaswinda, N. 2021. Analysis of rasch model for the validation of chemistry national exam instruments. *Jurnal Pendidikan Sains Indonesia (Indonesian Journal of Science Education)*, 9(3):329–345. <https://doi.org/10.24815/jpsi.v9i3.19618>
- Hebebcj, M.T. & Usta, E. 2022. The effects of integrated STEM education practices on problem solving skills, scientific creativity, and critical thinking dispositions.

- Participatory Educational Research*, 9(6):358–379. <https://doi.org/10.17275/per.22.143.9.6>
- Inayati, U. 2022. Konsep dan implementasi kurikulum merdeka pada pembelajaran abad 21 di SD/MI. *ICIE (Internasional Conference on Islamic Education)*, 2(8.5.2017):293–304. <http://proceeding.iainkudus.ac.id/index.php/ICIE>.
- Jeanne, H.S. & Prendergast, C.O. 2020. The assessment skills framework: a taxonomy of assessment knowledge, skills and attitudes authors. *Journal of Research and Practice in Assessment*, 15(1):1–25. <https://eric.ed.gov/?id=EJ1269869>.
- Khaldun, I., Hanum, L., & Utami, S.D. 2020. Pengembangan soal kimia higher order thinking skills berbasis komputer dengan wondershare quiz creator materi hidrolisis garam dan larutan penyangga. *Jurnal Pendidikan Sains Indonesia (Indonesian Journal of Science Education)*, 7(2):132–142. <https://doi.org/10.24815/jpsi.v7i2.14702>
- Kristanto, P.D. & Setiawan, P.G.F. 2020. Pengembangan soal hots (higher order thinking skills) terkait dengan konteks pedesaan. *Prisma: Prosiding Seminar Nasional Matematika*, 3:370–376. <http://journal.unnes.ac.id/sju/index.php/prisma/>
- Ndiung, S. & Jediut, M. 2020. Pengembangan instrumen tes hasil belajar matematika peserta didik sekolah dasar berorientasi pada berpikir tingkat tinggi. *Premiere Educandum: Jurnal Pendidikan Dasar dan Pembelajaran*, 10(1):94-99. <https://doi.org/10.25273/pe.v10i1.6274>
- Nugroho, A.N. & Airlan, A.S. 2020. Pengembangan instrumen penilaian keterampilan berpikir kritis pembelajaran IPA Kelas 4 SD. *Jurnal Ilmiah Pendidikan Profesi Guru*, 3(3):400-407. <https://dx.doi.org/10.23887/jippg.v3i3>.
- Nurmasyitah, P., Salim, A., Fransiska, I., Daris, K., & Suryani, K. 2023. *Implementation merdeka curriculum of learning to students' learning activities*, 3(1):39-44. <https://doi.org/10.30596/jcositte.v1i1.xxxx>
- Patonah, S., Nuvitalia, D., Saptaningrum, E., & Wuryandini, E. 2020. Inquiry learning tools to train science process skills at the junior high school level. *JPPS (Jurnal Penelitian Pendidikan Sains)*, 9(2):1812–1818. <https://doi.org/10.26740/jpps.v9n2.p1812-1818>
- Phil, S.M. 2021. *Pembelajaran dan Asesmen Kompetensi*. Badan Penelitian dan Pengembangan dan Perbukuan, Jakarta.
- Purnawanto, A.T. 2020. Perencanaan pembelajaran bermakna dan asesmen kurikulum merdeka. *Jurnal Ilmiah Pedagogy*, 20(1):75-94. <http://www.jurnal.staimuhblora.ac.id/index.php/pedagogy/article/view/116>
- Purniasari, L., Masykuri, M., & Ariani, S.R.D. 2021. Analisis butir soal ujian sekolah mata pelajaran kimia SMA N 1 kutowinangun tahun pelajaran 2019/2020 menggunakan model iteman dan rasch. *Jurnal Pendidikan Kimia*, 10(2):205–214. <https://jurnal.uns.ac.id/jpkim>.
- Puspita, K., Nazar, M., Hanum, L., & Reza, M. 2021. Pengembangan E-modul Praktikum Kimia Dasar Menggunakan Aplikasi Canva Design. *JUPI (Jurnal IPA & Pembelajaran*

- IPA), 5(2):151–161. <https://doi.org/10.24815/jipi.v5i2.20334>
- Ramadhani, R. & Fitri, Y. 2020. Validitas e-modul matematika berbasis epub3 menggunakan analisis rasch model. *Jurnal Gantang*, 5(2):95–111. <https://doi.org/10.31629/jg.v5i2.2535>.
- Rosnaeni, R. 2021. Karakteristik dan asesmen pembelajaran abad 21. *Jurnal Basicedu*, 5(5):4341–4350. <https://doi.org/10.31004/basicedu.v5i5.1548>.
- Setemen, K. 2018. Pengembangan dan pengujian validitas butir instrumen kecerdasan logis-matematis. *Jurnal Pendidikan Teknologi dan Kejuruan*, 15(2):178-187. <https://ejournal.undiksha.ac.id/index.php/JPTK/article/view/14197/9437>.
- Subayani, N.W. 2022. Implementasi STEM (science, technology, engineering, and mathematics) dalam kurikulum PGSD. *Didaktika: Jurnal Pemikiran Pendidikan*, 28(2):42-49. <https://doi.org/10.30587/didaktika.v28i2.4435>.
- Sumintono, B. & Widhiarso, W. 2014. *Aplikasi Model Rasch Untuk Penelitian Ilmu-Ilmu Sosial*. Trim Komunikata Pubishing House, Cimahi.
- Sumintono, B. & Widhiarso, W. 2015. *Aplikasi Pemodelan Rasch Pada Assessment Pendidikan*. Trim Komunikata Pubishing House, Cimahi
- Suparman. 2020. Menemukan karakteristik butir menggunakan quest. *Jurnal Komunkasi dan Pendidikan Islam*, 9(1):83–104. <https://doi.org/10.36668/jal.v9i1.134>.
- Sustiningsih, Utaminingsih, S., & Santoso. 2021. Development of science teaching materials based on stem: a needs analysis. *Journal of Physics: Conference Series*, 1823(1). <https://doi.org/10.1088/1742-6596/1823/1/012081>.
- Tursinawati, T. & Widodo, A. 2019. Pemahaman Nature of Science (NoS) Di Era Digital: Perspektif Dari Mahasiswa PGSD. *JUPI (Jurnal IPA & Pembelajaran IPA)*, 3(1):1–9. <https://doi.org/10.24815/jipi.v3i1.13294>
- Wang, J.C. & Wang, T.H. 2023. Learning effectiveness of energy education in junior high schools: Implementation of action research and the predict–observe–explain model to STEM course. *Heliyon*, 9(3):e14058. <https://doi.org/10.1016/j.heliyon.2023.e14058>.
- Widarta, F.O. & Artika, W. 2021. Analisis bentuk stimulus, dimensi kognitif, dan karakteristik HOTS pada instrumen evaluasi mata pelajaran IPA karya guru. *JUPI (Jurnal IPA & Pembelajaran IPA)*, 5(3):197–208. <https://doi.org/10.24815/jipi.v5i3.21429>
- Wulandari, T., Ramli, M., & Muzzazinah, M. 2022. Analisis butir soal dynamic assessment untuk mengukur pemahaman konsep klasifikasi tumbuhan pada mahasiswa. *Jurnal Pendidikan Sains Indonesia (Indonesian Journal of Science Education)*, 10(1):191–201. <https://doi.org/10.24815/jpsi.v10i1.22082>
- Yuliandini, N., Hamdu, G., & Respati, R. 2019. Pengembangan soal tes berbasis higher order thinking skill (HOTS) taksonomi bloom revisi di sekolah dasar. *Pedadidaktika: Jurnal Ilmiah Pendidikan Guru Sekolah Dasar*, 6(1):37–46. <http://ejournal.upi.edu/index.php/pedadidaktita/index>.

Zainal, N.F. 2020. Pengukuran, assessment dan evaluasi dalam pembelajaran matematika. *Laplace : Jurnal Pendidikan Matematika*, 3(1):8–26. <https://doi.org/10.31537/laplace.v3i1.310>.