

Rasch Model Analysis of the Quality of a Problem-Solving Skill Assessment Instrument on the Topic of Alternative Energy

Anggi Zazan Luvita¹, Abu Zainuddin²

^{1,2}State University of Surabaya , Indonesia

Article Info

Article history:

Accepted: 29 July 2026

Publish: 04 Agustus 2026

Keywords:

Rasch Model;

Instrument Quality;

Digital Assessment;

Problem-Solving Skills;

Socio-Scientific Issue;

Alternative Energy;

Abstract

Problem-solving skill is a crucial 21st-century competency that requires a high-quality assessment instrument for accurate measurement. This study aimed to analyze the quality of a Google Forms-based problem-solving assessment instrument—incorporating TikTok videos on Socio-Scientific Issues (SSI) regarding alternative energy—using the Rasch Model. As part of a research and development project, this study focused on analyzing instrument quality based on limited empirical testing involving 33 eleventh-grade students at SMA Negeri 1 Menganti. Instrument quality was analyzed using Winsteps software based on parameters such as item fit, reliability, separation index, and item difficulty. The results demonstrated that the instrument met high validity criteria, as confirmed by three experts who assigned a mode score of 4 across content, construct, and language aspects; the Percentage of Agreement (PoA) further indicated a very high level of inter-validator consensus. Rasch Model analysis revealed that all test items met item fit criteria based on Outfit MNSQ, Outfit ZSTD, and Pt Measure Corr values. The instrument also exhibited good reliability, with person separation and item separation values of 2.26 and 2.81, respectively, demonstrating its capability to differentiate between student ability levels and item difficulty levels. Item difficulty analysis identified four categories: 4 items were very easy, 3 were easy, 6 were difficult, and 3 were very difficult. Consequently, the developed assessment instrument meets quality standards and is suitable for measuring students' problem-solving skills.

This is an open access article under the [Lisensi Creative Commons Atribusi-BerbagiSerupa 4.0 Internasional](https://creativecommons.org/licenses/by-sa/4.0/)



Corresponding Author:

Anggi Zazan Luvita

State University of Surabaya, Indonesia

Email Coresspondent: anggizazandoc@gmail.com

1. INTRODUCTION

The rapid advancement of science and technology in the 21st century has brought about changes in various aspects of life, including the field of education [1]. This situation requires students not only to master knowledge but also to possess the various skills needed to face global challenges. One of the skills highlighted in the Partnership for 21st Century Skills (P21) framework is problem-solving, which falls under the category of learning and innovation skills. Problem-solving skills contribute significantly to developing the ability to address and resolve issues [2]. These skills enable students to identify problems, analyze information, plan strategies, and evaluate the resulting solutions [3]. Therefore, it is essential to both develop and accurately measure problem-solving skills through high-quality assessments that provide precise information regarding students' capabilities.

Students' problem-solving skills cannot be optimally measured solely through assessments focused on conceptual mastery or rote memorization. Measuring these skills requires assessment instruments capable of exploring students' abilities to understand problems, formulate solution strategies, execute solutions, and evaluate the results obtained [4]. With technological advancements, assessment practices have transformed into digital assessments, which offer greater flexibility and efficiency while facilitating the presentation of diverse learning stimuli. A systematic literature review indicates that digital assessments are more effective at measuring problem-solving skills; they do not merely focus on final outcomes but also enable the assessment of thought processes, solution strategies, and student reflection during the problem-solving journey [5]. Google Forms is a widely utilized platform for digital assessment due to its accessibility, support for various question types, and ease of managing and collecting student responses [6]. Furthermore, Google Forms can be integrated with learning media such as videos, making the presentation of contextual stimuli more engaging and effectively supporting the assessment of students' problem-solving capabilities.

The use of visual stimuli in test instruments—such as videos, images, graphs, animations, and interactive simulations featuring feedback and real-life scenarios—has been shown to stimulate students' higher-order thinking skills [7]. Consequently, utilizing video as a stimulus in digital assessments can present more contextual problems, thereby encouraging students to apply problem-solving skills. One medium with such potential is the TikTok video designed using the Socio-Scientific Issue (SSI) approach. The SSI approach can foster greater student engagement in the learning process [8]. By presenting scientific issues relevant to daily life, SSI helps students connect learned concepts with real-world problems. Therefore, integrating SSI-based TikTok videos into digital assessment instruments is expected to provide a more authentic context for measuring students' problem-solving skills.

Despite the growing use of digital assessments, the quality of the instrument remains a crucial factor in ensuring accurate and reliable measurement results. A high-quality instrument must accurately measure the intended construct, yield consistent measurements, and possess sound item characteristics [9]. Therefore, an empirical analysis of instrument quality is necessary to evaluate the characteristics of each item before the instrument is utilized for assessment purposes. The Rasch Model is a widely used approach for this purpose, as it enables the analysis of instrument quality based on parameters such as item fit, reliability, separation index, and item difficulty, thereby providing comprehensive information regarding the quality of the developed instrument.

Several studies have demonstrated the potential of utilizing technology and contextual approaches in both assessment and instruction. [10] reported that Google Forms (G-Form) serves as a highly practical and effective assessment medium, streamlining the entire process—from instrument creation, distribution, and administration to the processing of assessment results. Meanwhile, [11] showed that incorporating Socio-Scientific Issues (SSI) into the topic of alternative energy facilitates the analysis of students' problem-solving skills. Additionally, [12] employed the Rasch Model to analyze the quality of a Higher-Order Thinking Skills (HOTS) test instrument on renewable energy, examining parameters such as item fit, reliability, and difficulty level. Their findings indicated that the Rasch Model provides comprehensive information regarding instrument characteristics. However, these studies have largely examined these aspects in isolation: research on G-Form focuses on its utility as a digital assessment platform; SSI-related research emphasizes its application in the learning

process to foster student skills; and research utilizing the Rasch Model focuses on evaluating instrument quality. To date, there is limited research that integrates G-Form, SSI-based TikTok videos (as contextual stimuli), and Rasch Model-based quality analysis into a unified framework for measuring problem-solving skills in the context of alternative energy. Therefore, this study aims to analyze the quality of a G-Form-based digital assessment instrument—supported by SSI-themed TikTok videos—using the Rasch Model, thereby generating empirical evidence regarding its suitability as a tool for measuring problem-solving skills.

2. METHOD

This study employs a Research and Development (R&D) method focused on analyzing the quality of a Google Forms-based assessment instrument—supplemented by socio-scientific issue (SSI) videos from TikTok—designed to measure students' problem-solving skills regarding alternative energy topics. Instrument quality analysis was conducted through expert judgment and empirical analysis using the Rasch Model. The study was carried out at SMA Negeri 1 Menganti, Gresik. The test instrument analyzed consisted of four branched essay items, with each item comprising four sub-questions. The research process began with the development of the Google Forms-based assessment instrument incorporating TikTok-based SSI videos. Subsequently, the developed instrument underwent validation by three experts to assess the appropriateness of its content, construct, and language. Revisions were made based on expert feedback, a trial run with 33 students, and the instrument quality analysis. The instrument's logical validity was analyzed based on expert ratings, utilizing the mode value to determine the feasibility level of each aspect and the Percentage of Agreement (PoA) to measure the level of consensus among the experts. The PoA calculation follows Borich (1994) using the following equation.

$$\text{Percentage of Agreement} = 100\% \times \left(1 - \frac{A - B}{A + B}\right)$$

Furthermore, the quality of the instrument was empirically analyzed using the Rasch Model with the aid of Winsteps software, based on parameters such as item fit, reliability, item separation, and item difficulty.

According to [13], item fit is considered valid based on three criteria:

- (1) The acceptable Outfit MNSQ (Mean Square) value is $0.5 < \text{MNSQ} < 1.5$
- (2) The acceptable Outfit ZSTD (Z-Standard) value is: $-2.0 < \text{ZSTD} < +2.0$
- (3) Pt Measure Corr (Point Measure Correlation) value: $0.4 < \text{Point Measure Corr} < 0.85$

Furthermore, reliability is an indicator used to describe the consistency of measurement results when an instrument is used repeatedly under relatively similar conditions. [13] state that in analyses using the Rasch Model, reliability values are determined based on the summary statistics table, specifically the person reliability, item reliability, and Cronbach's alpha values. The criteria for each parameter are presented in the following table.

Table 1 Person/item reliability criteria in Rasch modeling

Nilai Person / Item Reliability	Interpretation
>0,94	Exceptional
0,91 – 0,94	Excellent
0,81 – 0,90	Good

0,67 – 0,80	Adequate
< 0,67	Weak

[13]

Table 2 The interpretation of reliability test results is based on the Cronbach's alpha value.

Nilai <i>Alpa Cronbach</i>	Interpretasi
$\alpha > 0,8$	Excellent
$0,7 < \alpha \leq 0,8$	Good
$0,6 < \alpha \leq 0,7$	Fair
$0,5 < \alpha \leq 0,6$	Poor
$\alpha < 0,5$	Very Poor

[13]

The item separation value is determined based on the summary statistics table, which indicates the extent to which the instrument can group test items according to their difficulty levels. A high separation value signifies superior instrument quality, as the instrument is capable of more accurately distinguishing between learner characteristics and item difficulty levels. The number of item difficulty groups is determined based on the separation value using the strata (H) equation, as follows.

$$H = \frac{[(4 \times \text{separation}) + 1]}{3}$$

Once the number of strata is determined, item difficulty categories are established based on item difficulty values (item measures) expressed in logits. The range of logit values is divided into intervals corresponding to the number of strata derived from item separation, allowing each item to be classified into a specific difficulty category.

3. RESULTS AND DISCUSSION

Instrument quality is a crucial factor determining the accuracy of measurement results. Therefore, a TikTok video-based, socio-scientific issue (SSI) assessment instrument—designed to measure students' problem-solving skills regarding alternative energy—was analyzed for both logical and empirical validity. Logical validity was established through evaluations by three expert validators: two physics education lecturers and one high school physics teacher. These validators were selected to ensure a comprehensive assessment covering scientific content, instrument development, and alignment with actual classroom implementation. The aspects evaluated included content, construct, and language. The results of the logical validity assessment are presented in Figure 1.

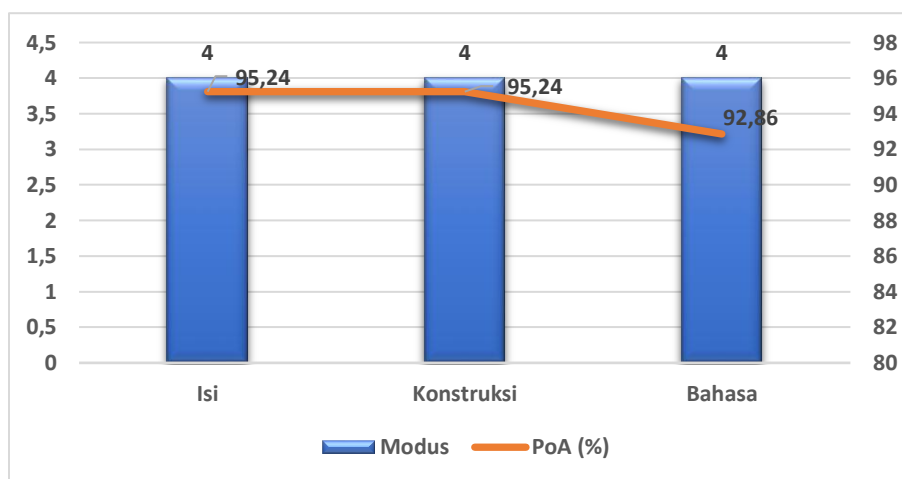


Figure 1 Diagram of validation results for the problem-solving skills instrument

Based on Figure 1, the assessment of content validity aims to ensure that the instrument aligns with learning outcomes (CP), learning objectives, and the specific problem-solving skill indicators being measured. Evaluations across the six aspects of content validity yielded a mode score of 4, placing the instrument in the "highly valid" category. These results indicate that each test item accurately represents the intended competencies and demonstrates a strong connection to the problem-solving skill indicators related to the topic of alternative energy. Consequently, the instrument has the potential to generate data that truly reflects students' abilities in accordance with the measured construct. Furthermore, the average Percentage of Agreement (PoA) of 95.24% indicates a very high level of consensus among validators, suggesting consistent evaluation of the instrument's content quality. This high level of agreement reinforces the conclusion that the instrument meets content validity requirements and is suitable for proceeding to the empirical testing stage. These findings align with research by [14], which states that an agreement percentage of 91.25%—or any figure exceeding 75%—can be considered reliable.

Construct validity is used to assess the alignment between the instrument's structure and the characteristics of the assessment being developed. Evaluations across four aspects yielded a mode score of 4 and an average Percentage of Agreement (PoA) of 95.24%, placing the instrument in the "highly valid" category. These results indicate that the instrument's components—such as instructions, the scoring system, the assessment rubric, and item construction—were systematically designed to optimally support the measurement of problem-solving skills. A well-structured instrument also enhances the consistency of assessment administration and minimizes potential measurement errors arising from an unclear format. These findings align with research by [15], which states that construct validity is demonstrated through the alignment of the instrument's constituent components, including instructions, assessment criteria, and scoring rubrics.

Validity regarding the linguistic aspect was assessed based on the clarity, precision, and accessibility of the language employed in the instrument. The assessment across four sub-aspects yielded a mode score of 4 and a mean Percentage of Agreement (PoA) of 92.86%, placing it in the "highly valid" category. These results indicate a high level of consensus among validators that the language used in the instrument adheres to Indonesian language standards, is easily understood by students, and is free from ambiguity. The use of clear, communicative language is a crucial factor, as it minimizes student errors in comprehending instructions or the context of the questions;

consequently, the resulting responses more accurately reflect problem-solving abilities rather than linguistic interpretation skills. Thus, these findings regarding linguistic validity further substantiate the instrument's suitability as a tool for measuring problem-solving skills in the context of alternative energy.

After being deemed suitable based on expert validation, the instrument was administered to students to gather empirical data. Data obtained from the trial involving 33 students at SMA Negeri 1 Menganti were subsequently analyzed using the Rasch Model to evaluate the instrument's quality based on its psychometric characteristics. The analysis covered item fit, reliability, separation, and item difficulty.

Table 3 Analysis of item fit

Question	Item Code	Outfit MNSQ	Outfit ZSTD	Pt Measure Corr	Category
1	BS1a	1,29	1,2	0,54	Item fit
	BS1b	0,77	-1,0	0,62	Item fit
	BS1c	1,04	0,3	0,47	Item fit
	BS1d	1,47	1,8	0,40	Item fit
2	BS2a	0,96	-0,1	0,71	Item fit
	BS2b	0,78	-0,9	0,77	Item fit
	BS2c	0,61	-1,9	0,69	Item fit
	BS2d	1,36	1,5	0,44	Item fit
3	BS3a	0,72	-1,3	0,52	Item fit
	BS3b	0,96	-0,1	0,69	Item fit
	BS3c	1,01	0,1	0,63	Item fit
	BS3d	0,83	-0,7	0,56	Item fit
4	BS4a	1,06	0,3	0,52	Item fit
	BS4b	0,90	-0,4	0,57	Item fit
	BS4c	0,97	0,0	0,58	Item fit
	BS4d	1,22	0,9	0,46	Item fit

Based on Table 3, the instrument quality analysis results indicate that all test items meet the item fit criteria for the Rasch Model. The instrument consists of four problems, each developed into four branches corresponding to the problem-solving stages of the model [16]. The code "BS" stands for "test item," while "a" denotes the stage of understanding the problem, "b" denotes devising a plan, "c" denotes carrying out the plan, and "d" denotes looking back. All items exhibit Outfit MNSQ, Outfit ZSTD, and Pt Measure Corr values within the ranges established by [13]; consequently, all items are empirically valid. These results demonstrate that each item functions in alignment with the problem-solving skill construct being measured and does not exhibit measurement deviation (misfit). Thus, student responses to each item are consistent with Rasch Model expectations, enabling the instrument to yield more accurate and reliable measurements. Furthermore, meeting the item fit criteria across all problem-solving stages indicates that each indicator adapted from the Polya model is well-represented within the instrument, making it suitable for empirically measuring students' problem-solving skills.

Table 4 Analysis of reliability

Aspect	Value	Category
<i>Person Reliability</i>	0,84	Good
<i>Item Reliability</i>	0,89	Good
<i>Cronbach's Alpha</i>	0,86	Very good

The next parameter for analyzing instrument quality is reliability. Based on Table 4, the person reliability value of 0.84 falls into the "good" category, indicating that student responses to the test items demonstrate a satisfactory level of consistency. In other words, the instrument yields relatively stable measurement results when used to assess the problem-solving skills of students with comparable characteristics. Meanwhile, the item reliability value of 0.89 also falls into the "good" category, indicating that the difficulty level of each item is consistently structured and of stable quality. This suggests that the variation in item difficulty is effectively maintained, allowing the instrument to adequately distinguish the characteristics of each item. Furthermore, the Cronbach's Alpha value of 0.86 falls into the "very good" category, reflecting high internal consistency between student abilities and item characteristics. These high reliability values demonstrate that the instrument not only produces consistent measurements but also possesses a high level of trustworthiness as a tool for assessing problem-solving skills. Consequently, the developed instrument has the potential to generate accurate and reliable data when utilized in the assessment process.

Table 5 Analysis of separation index

Aspect	Value	Strata (H)	Group
<i>Person Separation</i>	2,26	3,34	3
<i>Item Separation</i>	2,81	4,08	4

Based on Table 5, a person separation value of 2.26 was obtained, indicating that the instrument effectively distinguishes between students based on their problem-solving skill levels. The calculation of the number of strata (H) yielded a value of 3.34, which was rounded to three ability strata. These results indicate that the instrument can categorize students into three distinct ability levels, enabling teachers to more objectively identify students with high, medium, or low problem-solving capabilities. This ability to differentiate between student proficiency levels is significant, as it allows assessment results to serve as a basis for designing instructional follow-up tailored to students' needs. Furthermore, an item separation value of 2.81 indicates that the instrument possesses a strong ability to differentiate between the difficulty levels of the test items. Based on the calculation of the number of strata (H), a value of 4.08 was obtained, which rounds to four strata of difficulty. These results demonstrate that the developed items exhibit an adequate range of difficulty levels, spanning from very easy to very difficult categories. This variation indicates that the instrument is not dominated by items of a single difficulty level, thereby enabling a more comprehensive assessment of students' problem-solving skills across various ability levels. Consequently, the diversity in item difficulty supports measurement accuracy and enhances the instrument's quality as an assessment tool.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ ZSTD	OUTFIT MNSQ ZSTD	PT-MEASURE CORR. EXP.	EXACT MATCH OBS% EXP%	Item
3	74	33	1.31	.28	1.02 .2	1.04 .3	.47 .56	60.6 56.7	BS1c
16	75	33	1.23	.28	1.20 .9	1.22 .9	.46 .56	36.4 56.2	BS4d
12	76	33	1.15	.28	.84 -.6	.83 -.7	.56 .56	54.5 56.5	BS3d
2	83	33	.60	.28	.77 -1.0	.77 -1.0	.62 .57	66.7 57.6	BS1b
10	84	33	.53	.28	.95 -.1	.96 -.1	.69 .57	60.6 57.6	BS3b
4	85	33	.45	.28	1.47 1.8	1.47 1.8	.40 .57	57.6 57.5	BS1d
6	87	33	.29	.28	.77 -1.0	.78 -.9	.77 .57	60.6 57.2	BS2b
11	87	33	.29	.28	1.02 .2	1.01 .1	.63 .57	57.6 57.2	BS3c
8	88	33	.22	.28	1.36 1.5	1.36 1.5	.44 .57	45.5 57.0	BS2d
15	93	33	-.17	.28	.98 .0	.97 .0	.58 .58	60.6 56.6	BS4c
7	96	33	-.40	.28	.61 -1.9	.61 -1.9	.69 .58	63.6 56.5	BS2c
14	96	33	-.40	.28	.90 -.4	.90 -.4	.57 .58	57.6 56.5	BS4b
5	103	33	-.97	.29	.98 .0	.96 -.1	.71 .57	54.5 56.9	BS2a
9	107	33	-1.31	.30	.75 -1.1	.72 -1.3	.52 .57	75.8 58.0	BS3a
13	107	33	-1.31	.30	1.09 .5	1.06 .3	.52 .57	72.7 58.0	BS4a
1	109	33	-1.49	.30	1.38 1.5	1.29 1.2	.54 .56	57.6 59.8	BS1a
MEAN	90.6	33.0	.00	.28	1.01 .0	1.00 .0		58.9 57.2	
S.D.	11.2	.0	.89	.01	.24 1.0	.23 1.0		9.0 .8	

Figure 2 Output table item measure

Figure 2 presents a mean value of 0.00 logit and a standard deviation (SD) of 0.89 logit. The mean value serves as the reference point, while the SD value is used to determine the boundaries for each difficulty level category. Based on these values, item difficulty is classified into four categories, as shown in Table 6.

Table 6 Category of difficulty level of questions

Kriteria	Measure	Item Code	Interpretation
$Measure\ logit > Mean + SD$	$> 0,89$	BS1c; BS4d; BS3d	Very difficult
$Mean < Measure \leq Mean + SD$	0,00 s.d 0,89	BS1b; BS3b; BS1d; BS2b; BS3c; BS2d	Difficult
$Mean - SD \leq Measure < Mean$	-0,89 s.d 0,99	BS4c, BS2c, BS4b	Easy
$Measure\ logit < Mean - SD$	$< -0,89$	BS2a; BS3a; BS4a; BS1a	Very easy

Based on Table 6, the developed test items exhibit a good range of difficulty levels, comprising 4 items categorized as very easy, 3 as easy, 6 as difficult, and 3 as very difficult. This variation indicates that the instrument is not dominated by items of any single difficulty level, thereby enabling a more comprehensive assessment of students' problem-solving skills across various ability levels. Such diversity in difficulty also enhances measurement accuracy by allowing the instrument to differentiate between students with low, moderate, and high levels of ability.

The analysis results also indicate that all test items coded 'a'—representing the "understand the problem" stage—fell into the "very easy" category. This finding suggests that the majority of students were able to identify information, grasp the problem context, and determine both the given data and the specific questions asked. Conversely, items classified as "difficult" or "very difficult" were predominantly those coded 'b', 'c', and 'd', which correspond to the stages of devising a plan, carrying out the plan, and looking back. This demonstrates that while students were generally capable of understanding the problems, they struggled with formulating solution strategies, applying concepts to solve the problems, and re-evaluating the accuracy of their solutions. These findings align with the

nature of problem-solving skills, which require higher-order thinking abilities during the later stages; thus, the developed instrument effectively identifies specific areas of the problem-solving process that require improvement.

Overall, the research findings indicate that the digital assessment instrument—based on Google Forms and incorporating Socio-Scientific Issue (SSI) TikTok videos—meets quality criteria from both theoretical and empirical perspectives. Theoretical validity was confirmed by validator assessments, showing the instrument to be appropriate in terms of content, construct, and language. Meanwhile, empirical validity—established through Rasch Model analysis—demonstrated that all items met fit criteria, exhibited good reliability and varied difficulty levels, and effectively distinguished the difficulty characteristics of each item. These findings suggest that the instrument is not only suitable as a measurement tool but is also capable of generating accurate and consistent data for identifying students' problem-solving skills regarding alternative energy topics.

The results not only demonstrate that the developed instrument meets high-quality standards but also offer practical implications for teachers and researchers in physics education. For teachers, this Google Forms-based digital assessment instrument—utilizing SSI TikTok videos—serves as an alternative method for measuring students' problem-solving skills through contextual issues relevant to daily life. Integrating Google Forms with SSI-based TikTok videos facilitates the assessment process, streamlining everything from stimulus presentation and automated response collection to efficient result processing. For researchers, this instrument can serve as a reference for developing digital assessments covering other physics topics or for different educational levels. Furthermore, the Rasch Model analysis provides empirical evidence regarding the instrument's quality, offering a benchmark for future instrument development research.

Nevertheless, this study has certain limitations that should be considered when interpreting the results. The empirical testing of the instrument involved only 33 students from a single school—SMA Negeri 1 Menganti—meaning the findings cannot yet be fully generalized to a broader population. Furthermore, the developed instrument focused exclusively on the topic of alternative energy; consequently, applying it to other physics topics would require further adaptation and validation. The study also concentrated on analyzing the instrument's quality based on logical validity and psychometric characteristics using the Rasch Model, without evaluating its effectiveness in enhancing students' problem-solving skills during the learning process. Therefore, future research should involve a larger and more diverse group of respondents, develop instruments for other science or physics topics, and examine their effectiveness in supporting learning and assessment processes.

Despite these limitations, the study's findings provide empirical evidence that the Google Forms-based digital assessment instrument—incorporating TikTok videos centered on socio-scientific issues—demonstrates high quality based on both logical validity and Rasch Model analysis. Consequently, this instrument holds potential as an alternative digital assessment capable of measuring students' problem-solving skills more comprehensively by presenting contextual and authentic problems.

4. CONCLUSION

Based on the research findings and discussion, it can be concluded that the Google Forms-based assessment instrument—incorporating TikTok videos featuring socio-scientific issues to measure

students' problem-solving skills regarding alternative energy—meets the quality criteria for a sound instrument and is suitable for use. Validity assessments conducted by three validators indicate that the instrument meets validity criteria across content, construct, and language aspects, with a mode score of 4 for each, placing it in the "highly valid" category. Furthermore, the very high Percentage of Agreement (PoA) value demonstrates an excellent level of consensus among the validators. Empirical data analysis using Winsteps software also reveals that all developed test items meet fit criteria based on Outfit MNSQ, Outfit ZSTD, and Pt Measure Corr values. Additionally, the instrument demonstrates good reliability, enabling consistent measurement. Based on person separation and item separation values, the instrument effectively distinguishes between students' ability levels and the difficulty levels of the test items; the item separation value indicates that the items can be categorized into four strata of difficulty. Subsequent analysis of item difficulty identified four categories: 4 items classified as very easy, 3 as easy, 6 as difficult, and 3 as very difficult. These findings demonstrate that the developed instrument yields accurate and consistent measurements, offering a potential digital assessment alternative for measuring students' problem-solving skills in the context of alternative energy.

5. ACKNOWLEDGMENTS

The author wishes to express gratitude to the supervising lecturer for providing guidance, input, and mentorship throughout the preparation of this article. Thanks are also extended to the validators who evaluated and offered suggestions on the research instrument, ensuring its suitability for use. Appreciation is conveyed to the Physics teacher at SMA Negeri 1 Menganti for their support and cooperation during the research, as well as to all students who participated as respondents during the instrument's pilot testing and implementation. It is hoped that this article will contribute to the development of physics learning assessments.

6. BIBLIOGRAPHY

- [1] D. Mahrunnisya, "Keterampilan Pembelajar Di Abad Ke-21," *JUPENJI J. Pendidik. Jompa Indones.*, vol. 2, no. 1, pp. 101–109, 2023, doi: 10.57218/jupenji.vol2.iss1.598.
- [2] M. G. Jayadiningrat and E. K. Ati, "Peningkatan Keterampilan Memecahkan Masalah Melalui Model Pembelajaran Problem Based Learning (PBL) pada Mata Pelajaran Kimia," *J. Pendidik. Kim.*, vol. 2, no. 1, pp. 1–7, 2018, doi: 10.23887/jpk.v2i1.14133.
- [3] I. N. Karma, S. Istiningsih, Nurhasanah, and I. D. Hastuti, "Penerapan Penilaian Proses dan Hasil Belajar Berbasis Kurikulum 2013 Bagi Guru Sekolah Dasar Negeri Gugus 5 Kota Madya Mataram 2018/2019," *J. Pendidik. dan Pengabd. Masyarakat*, vol. 2, no. 2, 2019, doi: 10.29303/jppm.v2i2.1093.
- [4] A. G. C. Wicaksono and E. Korom, "Review of problem-solving measurement: An assessment developed in the Indonesian context," *Particip. Educ. Res.*, vol. 9, no. 1, pp. 116–136, 2022, doi: 10.17275/per.22.7.9.1.
- [5] R. Rahayu, H. Mazlina, and N. Azzahira, "Efektivitas Asesmen Digital dan Konvensional dalam Mengukur Kemampuan Pemecahan Masalah Sains: Systematic Literature Review," *Al- Ikhtibar J. Ilmu Pendidik.*, vol. 13, no. 1, pp. 111–123, 2026, doi: 10.32505/ikhtibar.v13i01.14462.
- [6] D. jaya Fitrohyati and R. Hikmawan, "Google Forms-Based Adaptive Assessment As A Digital Assessment Strategy To Examine Senior High School Students' Engagement And Cognitive Load," *JKTP J. Kaji. Teknol. Pendidik.*, vol. 9, no. 2, pp. 224–232, 2026, doi: 10.17977/um038v9i22026p224-232.

- [7] A. Barata, H. L. Mampouw, S. Mulyani, U. Kristen, S. Wacana, and S. M. P. N. Salatiga, “Peningkatan Epistemic Curiosity dan Kemampuan Pemecahan Masalah Matematika Siswa Kelas 7 berbantuan Video,” *J. Pendidik. Mat.*, vol. 3, no. 1, pp. 96–113, 2023, doi: 10.28918/circle.v3i1.358.
- [8] Siska, W. Triani, Y. Maryuningsih, and M. Ubaidillah, “Penerapan Pembelajaran Berbasis Socio Scientific Issues,” *EduSains J. Pendidik. Sains Mat.*, vol. 8, no. 1, pp. 22–33, 2020, doi: 10.23971/eds.v8i1.1490.
- [9] Z. Arifin, “Kriteria Instrumen dalam suatu Penelitian,” *J. THEOREMS (The Orig. Res. Math.*, vol. 2, no. 1, pp. 28–36, 2017.
- [10] W. Candrawati, L. N. Setiana, and O. P. Wardani, “Penggunaan Media Google Form sebagai Asesmen Diagnostik di SMA Negeri 2 Semarang Pendahuluan,” *DEIKTIS J. Pendidik. Bhs. dan Sastra*, vol. 5, no. 3, pp. 1550–1559, 2025, doi: 10.53769/deiktis.v5i3.1471.
- [11] D. P. Sahertian and S. N. Hidayati, “Analisis Keterampilan Pemecahan Masalah Siswa Berbantuan Artikel Socio-Scientific Issue pada Materi Energi Alternatif,” *PENSA E-JURNAL Pendidik. SAINS* |, vol. 10, no. 1, pp. 1–7, 2022, doi: 10.26740/pensa.v10i1.41128.
- [12] N. Rahmatillah and A. Zainuddin, “Evaluation of HOTS Test on Renewable Energy Materials through Rasch Model Analysis,” *J. Pijar MIPA*, vol. 20, no. 5, pp. 933–939, 2025, doi: 10.29303/jpm.v20i5.9458.
- [13] B. Sumintono and W. Widhiarso, “Aplikasi Model Rasch untuk Penelitian Ilmu-Ilmu Sosial,” no. 2014, 2015.
- [14] A. A. Wicaksono, T. Sunarti, and A. Zainuddin, “Validitas dan Reliabilitas Instrumen Penilaian Kinerja Praktikum Pada Materi Alat Optik Kelas XI SMA,” *IPF Inov. Pendidik. Fis.*, vol. 09, no. 03, pp. 325–331, 2020.
- [15] Y. Sherliyanti and M. N. R. Jauhariyah, “Validitas Instrumen Penilaian Berbasis Kelas untuk Mengukur Kompetensi Literasi Sains Peserta Didik pada Materi Pemanasan Global,” *IPF Inov. Pendidik. Fis. Pendidik. Fis.*, vol. 13, no. 3, pp. 168–176, 2024.
- [16] G. Polya, *How to Solve It: A New Aspect of Mathematical Method*. New Jersey: Pricenton Uniersity Press, 1973.