Chemined 14 (1) (2025)

Chemistry in Education

https://journal.unnes.ac.id/journals/chemined

Development of Two-Tier Multiple-Choice Diagnostic Test Instrument Based on Science Literacy on Chemical Equilibrium Material

Salsabil Zhafirah Oktaviani[1*)] and Harjono[1]

Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Negeri Semarang, Indonesia

ARTICLE INFO

Article history:
Received: February 2025
Accepted: June 2025
Published: July 2025

Kata Kunci:
Tes Diagnostik, Literasi Sains, TTMC, Kesetimbangan Kimia

Keywords:
Diagnostic Test; Science Literacy; TTMC, Chemical Equilibrium

DOI:
10.15294/chemined.v14i1.21813

ABSTRAK

Penelitian ini bertujuan untuk mengembangkan instrumen tes diagnostik Two-Tier Multiple Choice (TTMC) berbasis literasi sains untuk mengukur kemampuan literasi sains peserta didik pada materi kesetimbangan kimia. Penelitian ini merupakan penelitian pengembangan (Research and Development) dengan model 4D (Define, Design, Develop, Disseminate). Subjek penelitian berjumlah 59 peserta didik. Instrumen yang dikembangkan berupa tes diagnostik TTMC sebanyak 25 butir soal. Pengembangan instrumen dilakukan melalui tiga tahapan utama, yaitu: (1) penentuan ruang lingkup materi kesetimbangan kimia, (2) pengumpulan informasi terkait penggunaan instrumen TTMC, dan (3) penyusunan serta validasi instrumen TTMC. Analisis data menggunakan model Rasch yang mencakup uji validitas ahli, validitas butir soal, reliabilitas, tingkat kesukaran, daya beda, serta analisis profil literasi sains peserta didik. Hasil penelitian menunjukkan bahwa instrumen yang dikembangkan dinyatakan valid dengan skor rata-rata validasi sebesar 102 dari total skor 128 dan hasil Content Validity Index (CVI) sebesar 1. Koefisien reliabilitas instrumen pada uji coba skala kecil dan besar berturut-turut sebesar 0,73 dan 0,78, menunjukkan tingkat konsistensi yang baik. Instrumen ini juga memiliki daya beda yang baik karena mampu mengelompokkan kemampuan peserta didik ke dalam tiga kategori. Analisis profil literasi sains peserta didik menunjukkan kategori cukup dengan persentase rata-rata 46%, yang meliputi aspek konten (47%), konteks (47%), dan kompetensi (45%).

ABSTRACT

This study aims to develop a Two-Tier Multiple Choice (TTMC) diagnostic test instrument based on science literacy to assess students' science literacy skills on chemical equilibrium material. This research employed a Research and Development (R&D) design using the 4D model (Define, Design, Develop, Disseminate). The research subjects consisted of 59 students. The developed instrument was a TTMC diagnostic test comprising 25 items. The development process included three main stages: (1) determining the scope of chemical equilibrium material, (2) collecting information regarding the use of TTMC instruments, and (3) constructing and validating the TTMC diagnostic instrument. Data were analyzed using the Rasch model, which involved expert validation, item validation, reliability testing, difficulty level analysis, item discrimination testing, and students' science literacy profiling. The results indicated that the developed instrument is valid with an average expert validation score of 102 out of 128 and a Scale-Content Validity Index (S-CVI) of 1. The reliability coefficient for the instrument was 0.73 for the small-scale trial and 0.78 for the large-scale trial, indicating good consistency. Furthermore, the instrument demonstrated good discriminative ability by categorizing students into three ability groups. The science literacy profile analysis showed that students' science literacy was in the "sufficient" category with an average percentage of 46%, consisting of content aspect (47%), context aspect (47%), and competency aspect (45%).

Correspondence address:
Gedung D6 Lt.2 Jl Raya Sekaran Gunungpati, Semarang
E-mail: salsabilzhafirah8@gmail.com

p-ISSN 2252-6609
e-ISSN 2502-6852

INTRODUCTION

The current implementation of education in Indonesia aims to enhance students' academic skills and personal development. It denotes the Merdeka Curriculum tailored to meet the demands of 21st-century skills (Usmaedi, 2021). The Merdeka Curriculum's learning process is governed by the Process Standards outlined in Permendikbud No. 16 of 2022, encompassing the planning, implementation, and assessment phases (Kemdikbud RI, 2022). The fundamental elements of educational objectives include attitudes, knowledge, and skills. To achieve these goals, it is necessary to evaluate learning outcomes to determine the extent of student mastery over the material and offer constructive feedback for both students and educators to improve the quality of learning (Futhona, 2017). Assessment in education plays a crucial role in collecting information pertinent to the learning process and attaining specific objectives (Nasution, 2022). Within the framework of the Merdeka Curriculum, there are several assessment types, including diagnostic, formative, and summative assessments. Diagnostic assessments are designed to ascertain students' preliminary understanding before starting learning, whereas formative assessments are conducted throughout the learning process to evaluate student progress and offer constructive feedback. Summative assessment is conducted at the end of the learning process to assess overall achievement (Muktamar et al., 2024). Yuliaristiawan (2023) noted that diagnostic tests may be administered at the end of learning to assess students' comprehension of particular content and offer feedback for educators to evaluate the efficacy of their instructional strategies. According to research by Soesilo et al. (2024), diagnostic tests administered at the end of learning effectively identify mastered and unmastered materials while guiding teachers in their subsequent planning.

Attaining 21st-century skills requires science literacy-based learning, enabling students to cultivate essential life skills for addressing contemporary challenges through critical, collaborative, and scientific thinking (Kasse & Atmojo, 2022). Science literacy-based learning equips students to comprehend natural phenomena and human activities, evaluate and design scientific research, and interpret scientific data and evidence. However, science education in Indonesia encounters considerable challenges, notably the low levels of students' science literacy compared to developed nations (Lestari et al., 2025). According to the 2018 Program for International Student Assessment (PISA), Indonesia ranked 71st among 79 participating countries, with an average score of 396 (OECD, 2019). Haryani and Prasetya (2021) found that 66% of Indonesian students scored below the minimum threshold for science competencies. This indicates that scientific literacy is essential for equipping students to address the growing complexity of issues in knowledge and technology (Hidayati & Julianto, 2018). One solution to improve scientific literacy is to provide scientific literacy-based practice questions. Septiani et al. (2019) found that students exhibit low scientific literacy skills due to insufficient training in responding to scientific literacy

questions. The assessment process in educational institutions also plays a significant role in this context. Therefore, measuring scientific literacy is crucial for evaluating students' understanding of scientific concepts, which in turn improves the quality of education in Indonesia and enhances its international competitiveness (Barusa et al., 2024).

Science literacy in chemistry learning is essential, as chemistry examines the properties and transformations of substances and the laws governing these processes (Effendy, 2017). A concept that students frequently find challenging to understand is chemical equilibrium. Chemical equilibrium, which elucidates alterations in the equilibrium state of reactions at the submicroscopic level, often renders such changes difficult to observe visually (Kurniyaningsih & Yonata, 2019). The Merdeka Curriculum also presents a significant challenge in chemical equilibrium, as numerous students find it difficult to understand the dynamic characteristics of equilibrium reactions (Gultom & Muchtar, 2022). An insufficient grasp of chemical equilibrium impedes comprehension of other chemical substances, such as acids and bases, salt hydrolysis, and buffer solutions (Marfu'a & Resti, 2022). Observations at SMAN 12 Semarang revealed that students faced difficulties in understanding the material on buffer solutions, particularly regarding the concept of equilibrium reactions that explain the function of the solution. Furthermore, interviews conducted with chemistry teachers at SMA Negeri 12 Semarang indicate the need for an assessment instrument that comprehensively, practically, and objectively evaluates students' abilities and measures their science literacy skills. The instrument currently employed is a descriptive assessment lacking a scientific literacy foundation. Interviews with educators at SMA Negeri 1 Batang and MAN 1 Jepara suggest that multiple-choice test instruments effectively assess learning outcomes. However, there are limitations in assessing the reasoning behind the chosen answer options. Developing a two-tier multiple choice (TTMC) diagnostic test instrument based on scientific literacy is crucial for evaluating students' understanding of scientific concepts and science literacy skills. Ramdayani et al. (2023) illustrate that the TTMC instrument, based on scientific literacy, effectively evaluates students' understanding of concepts, especially in acid-base topics.

METHODS

This study aims to develop a two-tier multiple choice (TTMC) diagnostic test instrument based on scientific literacy regarding chemical equilibrium, employing a research and development (R&D) method with a quantitative approach. This method was chosen to generate a valid product and evaluate its effectiveness within the learning context. This study employs a quantitative approach utilizing the R&D development method and the 4-D development model (define, design, develop, disseminate) to create a TTMC diagnostic test instrument grounded in scientific literacy concerning chemical equilibrium material. The define stage encompasses the analysis of student needs, task analysis, conceptual

frameworks, and the formulation of learning objectives, all of which are integrated to ensure the instrument's relevance and accuracy. The instrument's validity is assessed through content and construct validation, confirming its alignment with the indicators of scientific literacy. Experts conduct content validity, which is subsequently analyzed statistically through CVI analysis. The Rasch model's validity is utilized to assess the accuracy of each question item concerning student abilities. This involves a probability analysis that accounts for question difficulty and respondent abilities and identifies misfit or outlier items. Reliability, defined as the consistency of results, is assessed using Cronbach's Alpha. A high value indicates that the instrument is deemed consistent and reliable. The Rasch model offers a linear scale that enhances data interpretation and objective assessment, enabling the prediction of absent data and independent comparison of research variables.

RESULT AND DISCUSSION

This study aims to develop a TTMC diagnostic test instrument grounded in scientific literacy concerning chemical equilibrium concepts. This product is presented in a paper-based format. This instrument serves as an evaluation tool to assess students' scientific literacy regarding chemical equilibrium concepts after studying the material. This study employs a Research and Development (RnD) approach with a 4D model based on the stages established by Thiagarajan et al. (1974), which include definition, design, development, and dissemination. The development of a TTMC diagnostic test instrument based on scientific literacy regarding chemical equilibrium emphasizes content, context, and competence to evaluate students' scientific literacy skills in grade XI, as detailed below.

The expert validation was performed by UNNES chemistry lecturers and chemistry teachers from SMAN 12 Semarang. It was performed to verify that the developed TTMC diagnostic test instrument exhibited strong content validity and was suitable before the trial. The validation sheet consists of material aspects, question construction aspects, language aspects, and expert comments. The material aspect pertains to the interconnection among learning objectives, question indicators, and scientific literacy. The construction includes criteria for diagnostic test questions, primary questions, discourse within questions, images accompanying questions, and answer choices categorized into tier 1 and tier 2. The language aspect encompasses both the language employed, the spelling utilized and the comments from validators. The findings indicate that the TTMC instrument developed is valid for research purposes.

The results regarding the content validity of the TTMC diagnostic test indicate that the instrument effectively assesses students' scientific literacy skills. The validation results of the non-test instrument indicate its appropriateness for research applications. The evaluation of the test instrument by two

validators yielded an average score of 102 out of 1028. This score was analyzed using the Content Validity Index (CVI), resulting in a Scale-CVI (S-CVI) of 1, indicating a very high level of validity. Additionally, the non-test instrument received a score of 36 from two validators. During the validation phase, the validator recommended enhancing the developed diagnostic test instrument.

Upon the validation of the diagnostic test instrument, which was confirmed as valid, revisions were made following the recommendations and instructions provided by the validator. Subsequently, the instrument was administered to students for testing. A small-scale trial was conducted involving 27 students from class XI F.7 at SMAN 12 Semarang to evaluate the quality of the developed test instrument. The duration of the test was 90 minutes. The analysis of the scientific literacy test items indicated that item number 15 is inappropriate, exhibiting an MNSQ value of 1.84 and a ZSTD of 2.31. The diagnostic test equipment utilized in the small-scale trial demonstrated an adequate reliability value. The person reliability value was 0.69, the person reliability value was 0.73, and the Cronbach's Alpha coefficient was 0.82. The reliability of the TTMC diagnostic test equipment is outlined in Table 1 below.

Table 1. The Reliability of The TTMC Diagnostic Test Instrument Small Scale Test

Types of Reliability	Value	Category
Person Reliability	0.69	Enough
Item Reliability	0.71	Enough

An analysis of the difficulty level was performed to evaluate the capacity of the questions to categorize students' cognitive abilities. The difficulty level analysis of the questions utilized the Rasch model, incorporating the item measure feature. A high logit value signifies that the questions are of considerable difficulty. The results indicated that students identified 13 of the 25 questions as difficult. The reason for this is that students have not yet become acquainted with the diagnostic test instruments that have been developed. Table 2 presents the detailed results of the item measure analysis conducted in small-scale trials.

Table 2. Item Measure Analysis Result of TTMC Diagnostic Test Small Scale Test

Category	Logit Value	Question Number	Total questions
Very Difficult	Measure logit > SD logit	6,13,16,20	4
Difficult	0 ≤ Measure logit ≤ SD logit	11,18,22,19,25,3,12,21,23	9
Currently	–SD logit ≤ Measure logit ≤ 0	7,5,2,4,17,14	6
Easy	Measure logit < –SD logit	15,10,24,8,9,1	6

The analysis of discriminative power for the small-scale test yielded a *separation person* value of 1.51 and a *separation item* value of 1.64. A higher separation value indicates an improved quality of the test instrument. Thus, the developed test instrument can identify groups of question item levels alongside students' cognitive abilities. The analysis of the H person yielded a value of 2.3, which has been rounded

to 2. The results indicate that the *separation person* is adequate, as it effectively differentiates between two distinct groups of students. The H item value of 2.52 can be approximated to 3. The results indicate that the separation item is classified as good, as it effectively distinguishes among three categories of questions. The final small-scale test analysis involves the *Wright map* analysis. The distribution indicates that the questions S13, S16, S20, and S6 exhibit the highest levels of difficulty. Conversely, question S1 is classified as easy due to its lowest logit value. The right-side displays students identified by code 26LD who have achieved the highest level of ability, attaining a score of 89 out of 100. Meanwhile, 15LK demonstrates a low performance level, achieving a score of 12 out of 100.

The validity analysis of the TTMC diagnostic test instrument is conducted by evaluating the test items to determine their appropriateness based on established criteria for item suitability. The analysis of test item validity indicates that several questions fail to satisfy the MNSQ value criteria. Specifically, question 22, which pertains to the chemical equilibrium factor in the formation process of Cl_2, and question 3, which addresses the factors influencing the temperature and pressure of ammonia gas, are identified as non-compliant. The MNSQ values for these test items are +1.68 and +1.67, respectively. Both questions fall outside the acceptable MNSQ value range of +0.5 to +1.5. Both items are maintained and can be utilized in large-scale tests, as the ZSTD values for both meet the established criteria. The diagnostic test instrument employed in extensive trials demonstrates adequate reliability values. The values for item reliability and person reliability in large-scale trials fall within the sufficient category. The reliability value for the person is 0.80, the reliability value for the item is 0.78, and the Cronbach Alpha is calculated at 0.87. Table 3 illustrates the reliability of the TTMC diagnostic test instrument.

Table 3. The Reliability of The TTMC Diagnostic Test Instrument Large-Scale Test

Types of Reliability	Value	Category
Person Reliability	0.80	Enough
Item Reliability	0.78	Enough

The difficulty level of the test items was analyzed on a large-scale test. A high logit value signifies that the test items exhibit a high difficulty level. The difficulty level of the test allows for the identification of test items that students perceive as challenging. The findings indicated that students identified 13 of the 25 test items as difficult. The analysis results of the TTMC diagnostic test measure items, which have undergone extensive development and testing on a large scale, are presented in Table 4.

Table 4. Item Measure Analysis Result of TTMC Diagnostic Test Small Scale Test

Category	Logit Value	Question Number	Total questions
Very Difficult	Measure logit > SDlogit	6,16,20,12,13	5
Difficult	0 ≤ Measure logit ≤ SDlogit	18,11,4,5,17,21,22,24	8

Currently	–SDlogit ≤ Measure logit ≤ 0	3,23,25,2,14,9	6
Easy	Measure logit < –SDlogit	7,8,10,15,19,1	6

The large-scale test discrimination power analysis yielded a person separation value of 2.00 and an item separation value of 1.86. A greater separation value indicates an improved quality of the test instrument. The developed test instrument is capable of identifying groups of question item levels alongside students' cognitive abilities. The H-person analysis yielded a result of 3. The results indicate that the categorization of person separation is classified as good, as it effectively differentiates among three distinct groups of students. The H item value of 2.82 is subject to rounding, resulting in a final value of 3. The results indicate that the item separation is effective, as it successfully differentiates among three categories of questions. The final comprehensive test analysis evaluates the Wright map by examining both ends of the distribution. The distribution presented on the Wright map indicates the highest difficulty level, specifically in question S6, which pertains to the impact of adding V_2O_5 on the sulfur dioxide contact process reaction with oxygen, exhibiting a logit value of +2.87. The lowest logit value was recorded for number S1, with a logit value -2.20. The right side displays the student's proficiency level for codes 04PK and 20PK, indicating a score of 84 out of 100. In contrast, code 03LK received a score of 10 out of 100 for students demonstrating the lowest level of ability.

The validation results of the TTMC diagnostic test instrument indicate its effectiveness in measuring students' scientific literacy following the established indicators. The sentences and structure of the questions are modified to align with the cognitive level, facilitating the evaluation of comprehension regarding the concepts of chemical equilibrium and equilibrium constants. Enhancements to the questions' context facilitate assessing students' capacity to analyze factors influencing reaction equilibrium, including temperature and pressure. This revision ensures that the instrument accurately represents the concept's significance in both daily life and industry, thereby facilitating opportunities for students to implement their knowledge in practical scenarios. The discriminatory power analysis indicates that the TTMC diagnostic test instrument effectively assesses students' scientific literacy. Assessment of students' comprehension regarding the influence of chemical equilibrium in industrial applications, exemplified by the Haber-Bosch process for ammonia synthesis, can be conducted through questions that illustrate practical scenarios. The results of the analysis indicate that students with high abilities perform better in answering questions related to this topic. Questions regarding the equilibrium of carbon dioxide with water in the environment were also identified. The capacity of students to respond to these inquiries indicates their comprehension of the impact of equilibrium reactions on the ecosystem. Health instruments serve as tools for assessing students' comprehension of the implications of chemical equilibrium within the human body, particularly regarding regulating oxygen and carbon dioxide levels

in the bloodstream. The pattern of students' responses can indicate the level of their comprehension of this concept.

The evaluation of the scientific content aspect indicates a score of 47%, while the scientific context aspect also reflects a score of 47%. The competency aspect is recorded at 45%. The overall percentage of students' scientific literacy within the context, content, and competency aspects is 46%, indicating a sufficient level of proficiency. The outcome of various factors that were carried out indicated the lowest percentage in the competency aspect, which was recorded at 45%. The results of the scientific literacy test delivered to 32 students are presented and evaluated based on context, content, and competency dimensions. The results can be found in Appendix 26. Sutrisna's research (2021) demonstrated that students' low scientific literacy skills are influenced by several factors, including a low interest in reading, evaluation tools that do not effectively promote the development of scientific literacy, and insufficient teacher knowledge regarding scientific literacy. Furthermore, the results of low scientific literacy assessments suggest a deficiency in the significance of the science learning process (Rosdiana *et al*., 2018).

The initial indicator in the science literacy inquiry pertains to science content. The question assesses three science content areas: 1) Analyzing and explaining the concept of chemical equilibrium, types of reactions, and equilibrium constants. Additionally, the data processing results will be presented to ascertain the value of the chemical equilibrium constant; 2) Analyze factors that influence and determine the direction of chemical reaction equilibrium; 3) Analyze equilibrium reactions in industrial applications and everyday scenarios. The calculation results regarding the achievement of science literacy skills in the overall content aspect are illustrated in Figure 1, which displays the average score of students alongside the percentage of students who answered each question item correctly.

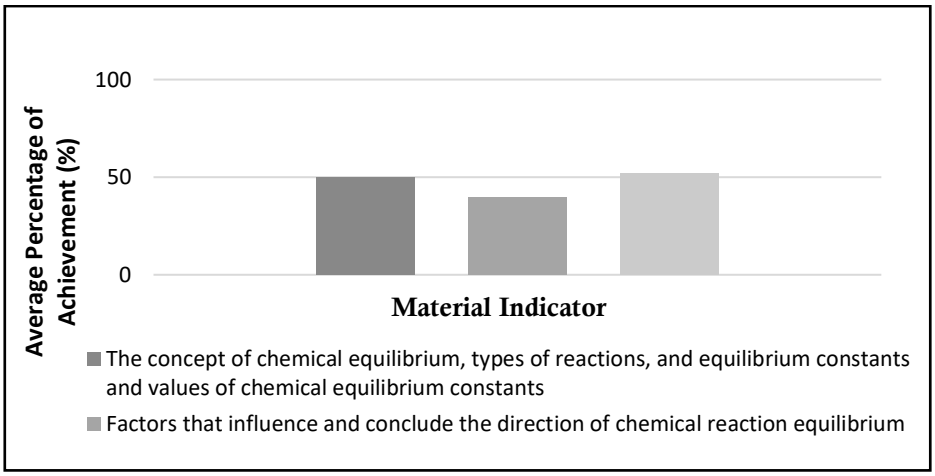

Figure 1. Average Score of Science Literacy Ability in Content Aspect

According to Figure 1, the overall content aspect of scientific literacy is measured at 47%, categorizing this achievement as "sufficient." The assessment of student performance regarding the concepts of chemical equilibrium, types of reactions, equilibrium constants, and the values of chemical equilibrium constants indicates an achievement level of 50%, categorized as "sufficient." The following indicator pertains to the factors that affect and determine the direction of chemical reaction equilibrium. The achievement level attained by students is 40%, categorized as "sufficient." The indicator measuring questions on equilibrium reactions in both industry and daily life shows an achievement level of 53%, categorized as "sufficient," the highest among the other two indicators. The results of the content aspect of scientific literacy demonstrate that students require enhancement in their comprehension of learning materials. While the emphasis on learning leans towards content, the mastery of content concepts among students remains insufficient. The insufficient level of students' scientific literacy in the content aspect suggests that they have not yet fully developed the ability to apply their knowledge in practical, everyday situations.

The second indicator related to scientific literacy pertains to the scientific context evaluated in the question, specifically within the industry field. The influence of chemical equilibrium on industrial processes, such as the Haber Bosch process for ammonia synthesis, and its implications for the environment. Equilibrium environmental reactions, such as the equilibrium of carbon dioxide with water, play a significant role in health. The influence of chemical equilibrium within the human body, such as the balance of oxygen and carbon dioxide in the bloodstream. The calculation results regarding the achievement of scientific literacy skills within the overall context aspect are illustrated in Figure 2. This figure displays students' average scores and the percentage of students who answered each question item correctly.

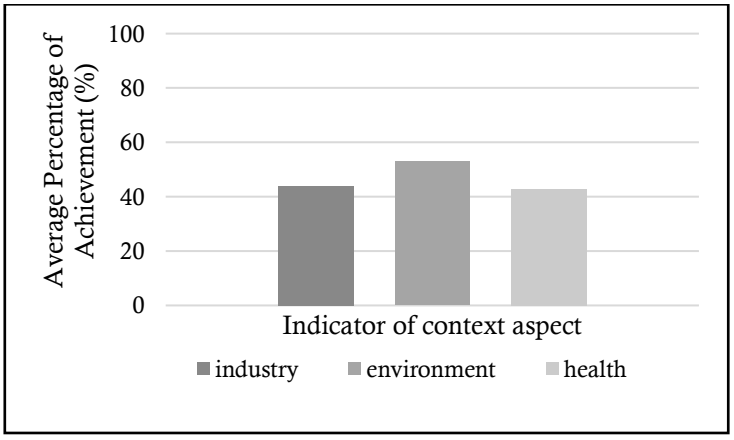

Figure 2. Average Score of Science Literacy Ability in Context Aspect

According to Figure 2, the overall achievement of scientific literacy in the context aspect is 47%, which falls within the "sufficient" category. The initial context indicator, the industry indicator, shows that students have achieved a score of 44%, which falls within the "sufficient" category. The subsequent indicator is the environmental indicator. The students achieved a score of 53%, with the "sufficient" category representing the highest level of achievement compared to the other two indicators. The health indicator shows that the achievement level attained by students is 43%, categorized as "sufficient." The limited capacity of students to comprehend the context results from an insufficient grasp of the concepts and processes being taught. Conversely, within the context aspect, students must demonstrate the ability to solve problems contextually.

The third indicator of scientific literacy pertains to the competency required in formulating questions, which includes 1) Explaining phenomena through a scientific lens, 2) Evaluating and designing scientific investigations, and 3) Interpreting scientific evidence and analyzing data. The calculation results regarding the attainment of scientific literacy skills within the overall competency aspect are illustrated in Figure 3. This figure displays the average score of students alongside the percentage of students who answered each question item correctly.

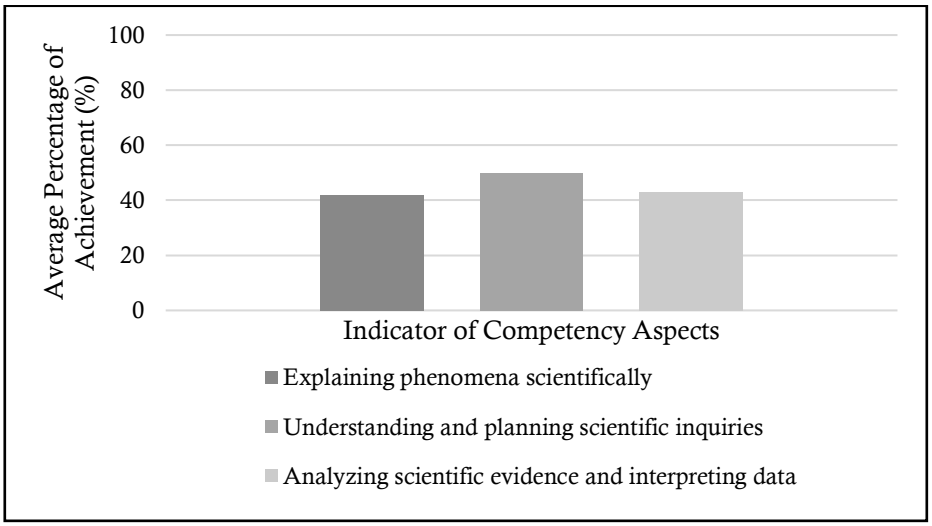

Figure 3. Average Score of Science Literacy Ability in Content Aspect

According to Figure 3, the overall competency aspect of scientific literacy is achieved at a rate of 45%, categorizing this achievement as "sufficient." The first competency indicator, which explains phenomena scientifically, shows that students scored 42%, falling within the "sufficient" category. The performance of this indicator is significantly lower than that of other indicators. The next indicator, which focuses on evaluating and designing scientific investigations, shows that students achieved a performance level of 50%. The highest achievement among the three indicators falls within the "sufficient" category.

The final indicator, which pertains to the interpretation of scientific evidence and data, shows that students scored 43%, categorized as "sufficient." The attainment of literacy skills within the competency framework, specifically regarding the indicator of explaining phenomena scientifically, remains significantly low. This is attributed to students' insufficient comprehension of the subject matter. The research conducted by Nofiana and Julianto (2018) indicates that a contributing factor to the low scientific literacy among students is their limited responsiveness to developments and issues in their environment, particularly those associated with natural phenomena, local and regional advantages, and surrounding environmental problems. Several factors contribute to low scientific literacy in Indonesia, as identified by the OECD in 2016. These factors include gender, economic conditions, and social influences.

CONCLUSION

The diagnostic test instrument comprises a test grid, test question sheets, and scoring guidelines. The final product is a TTMC diagnostic test instrument comprising 25 questions. The validator's assessment indicates that the TTMC diagnostic test instrument, which focuses on scientific literacy in chemical equilibrium, demonstrates good feasibility, achieving an average validation score of 102 out of a possible 128, categorizing it as valid. The score was analyzed using the Content Validity Index (CVI), with a Scale-level Content Validity Index (S-CVI) of 1, indicating consistency in the experts' responses. The item reliability test yielded a result of 0.78, categorizing it as sufficient after implementation with students. The developed TTMC diagnostic test instrument is intended for use. The analysis of students' scientific literacy abilities revealed that the achievement levels for content, process, and context were 47%, 47%, and 45%, respectively. The average overall percentage is 46%, categorizing it as sufficient. However, habituation is necessary to enhance scientific literacy, enabling students to become familiar with scientific literacy questions.

The suggestion from this study is that teachers and educational developers can consider using this science literacy-based TTMC instrument as an alternative to more in-depth evaluation, especially in science subjects that require strong conceptual understanding. The use of the Rasch model is also recommended in the development of other educational instruments to ensure objectivity and reliability of measurement results, so that the evaluation results more accurately reflect the abilities of students. Further research can be conducted to test this instrument on different materials or at various levels of education to expand the generalizability of the results. In addition, teachers are advised to utilize the results of this diagnostic instrument as a guide in designing more targeted learning interventions, especially in

overcoming learners' misconceptions and strengthening their science literacy skills to be better prepared to face the challenges of the 21st century learning era.

REFERENCE

Barusa, R. A., Rusilowati, A., & Ridlo, S. (2024). *Analisis Kebutuhan Pengembangan Instrumen Tes Penilaian Literasi Sains Berorientasi TIMSS Siswa SD Kelas V*. *12*(1), 68–85.

Bayrak, B. K. (2013). Using Two-Tier Test to Identify Primary Students' Conceptual Understanding and Alternative Conceptions in Acid Base. *Mevlana International Journal of Education*, *3*(2), 19–26.

Boone, W. J., Staver, J. ., & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences.* (springer).

Davis, L. L. (1992). Instrument review: Getting the most from a panel of experts. *Applied Nursing Research*, *5*(4), 194–197. https://doi.org/10.1016/S0897-1897(05)80008-4

Effendy. (2017). *Molekul Struktur, dan Sifat-Sifatnya*. Indonesian Academia Publishing.

Futhona, A. (2017). *Pengembangan Kumpulan Soal Pengayaan Kimia Berbasis Higher Order Thinking Skill (HOTS) Materi Asam Basa, Hidrolisis dan Larutan Penyangga*. Universitas Islam Negeri Sunan Kalijaga.

Gultom, E. R., & Muchtar, Z. (2022). Penerapan Model Pembelajaran Berbasis Proyek Pada Pembelajaran E-Learning Materi Kesetimbangan Kimia. *Educenter : Jurnal Ilmiah Pendidikan*, *1*(4), 327–334. https://doi.org/10.55904/educenter.v1i4.91

Gurel, D. K., Eryilmaz, A., & McDermott, L. C. (2015). A review and comparison of diagnostic instruments to identify students' misconceptions in science. *Eurasia Journal of Mathematics, Science and Technology Education*, *11*(5), 989–1008. https://doi.org/10.12973/eurasia.2015.1369a

Haryani, S., & Prasetya, A. T. (2021). *Desain Perangkat Pembelajaran Terintegrasi Kecakapan Abad 21*. DIVA press.

Hidayah, N., Rusilowati, A., & Masturi, M. (2019). Analisis Profil Kemampuan Literasi Sains Siswa SMP/MTs di Kabupaten Pati. *Phenomenon : Jurnal Pendidikan MIPA*, *9*(1), 36.

Hidayati, & Julianto. (2018). Penerapan Literasi Sains dalam Pembelajaran IPA di Sekolah Dasar untuk Meningkatkan Kemampuan Berfikir Kritis Siswa dalam Memecahkan Masalah. *Seminar Nasional Pendidikan Banjarmasin*.

Kasse, F., & Atmojo, I. R. W. (2022). Analisis Kecakapan Abad 21 Melalui Literasi Sains Pada Siswa Sekolah Dasar. *Jurnal Education and developmen*, *10*(1), 124–128.

Kemdikbud RI. (2022). Peraturan menteri pendidikan kebudayaan riset dan teknologi tentang standar proses pada pendidikan usia dini, jenjang pendidikan dasar dan jenjang pendidikan menengah. Peraturan Menteri Pendidikan Dan Kebudayaan Republik Indonesia Nomor 16 Tahun 2022 Tent (hal. 5–24).

Kurniyaningsih, B., & Yonata, B. (2019). LembarKegiatan Peserta DidikBerbasis Inkuiri Terbimbing Untuk Melatihkan Keterampilan Berpikir Kreatif Peserta Didik Pada Materi KesetimbanganKimia. *UNESA Journal of ChemicalEducation*, *8*(1), 75–81.

Lestari, D., Oktiani, R., Lutfianasari, U., & Imaduddin, M. (2025). Chemistry Teachers' Perspectives on the Implementation of Interdisciplinary Science Projects in Vocational Schools: Challenges and Strategic Solutions. JTK (Jurnal Tadris Kimiya), 10(1), 76–91. https://doi.org/10.15575/jtk.v10i1.37830.

Marfu'a, S., & Resti, T. A. (2022). Analisis Kesulitan Belajar Siswa Dalam Memahamimaterikesetimbangankimia. *Prosiding Seminar Nasional Pendidikan Kimia 2022*.

Muktamar, A., Ardianto, & Ariswanto. (2024). Optimalisasi Pembelajaran Melalui Implementasi Asesmen Pembelajaran pada Kurikulum Merdeka. *Journal of International Multidisciplinary Research*, *2*(4),10–18. https://journal.banjaresepacific.com/index.php/jimr/article/view/395%0Ahttps://journal.banjaresepacific.com/index.php/jimr/article/download/395/360

Myanda, A. A., Riezky, M. P., & Maridi, M. (2020). Development of Two-Tier Multiple-Choice Test to Assess Students' Conceptual Understanding on Respiratory System Material of 11th Grade of Senior High School. *International Journal of Science and Applied Science: Conference Series*, *4*(1), 44. https://doi.org/10.20961/ijsascs.v4i1.49457

Nalina, T. & Z.H.A. Rahim. 2007. The crude aqueous extract of Piper betle L. and its antibacterial effect towards Streptococcus mutans. *Jurnal of Biochem & Biotech,* 3(1):10-5

Nasution, S. W. (2022). Assesment Kurikulum Merdeka Belajar Di SekolahDasar. *Prosiding Pendidikan Dasar*, *1*(1), 135–142.

OECD. (2019b). *Take The Test: Sample Questions From Oecd"s Pisa Assessments*. OECD.

OECD. (2016). *PISA 2015 Assessment And Analytical Framework: Science, Reading, Mathematic And Financial Literacy*. OECD publishing.

Ramdayani, R., Savalas, L. R. T., & Idrus, S. W. Al. (2023). *Pengembangan intrumen evaluasi pembelajaran kimia two-tier multiple choice unyuk mengukur kemampuan literasi sains siswa kelas XI IPA MAN 2 mataram pada materi asam basa*.

Septiani, D., Widiyawati, Y., & Nurwahidah, I. (2019). Pengembangan Instrumen Tes Literasi Sains

Berbasis Pisa Pada Aspek Menjelaskan Fenomena Ilmiah Untuk Siswa Kelas Vii. *Science Education and Application Journal*, *1*(2), 46. https://doi.org/10.30736/seaj.v1i2.144

Soesilo, T. D., Kristin, F., & Windrawanto, Y. (2024). Pelaksanaan Diagnostik Kesulitan Belajar Peserta Didik oleh Guru SD di Kota Salatiga. *Scholaria: Jurnal Pendidikan dan Kebudayaan*, *14*(01), 59–67. https://doi.org/10.24246/j.js.2024.v14.i01.p59-67

Sumarni, W., Prasida, H. W., & Sumarti, S. S. (2017). Pengembangan Instrumen Penilaian Kemampuan Kognitif dan Afektif Berbasis Literasi Sains pada Materi Larutan Penyangga. *Seminar Nasional Alfa VII*.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan RASCH Pada Assessment Pendidikan*. C. T. Komunikata (ed.).

Thiagarajan, S., Semmel, D. ., & Semmel, M. I. (1974). *Instructional development for training teacher of exceptional children*. Indiana University.

Thummathong R, T. K. (2016). Construction of a chemical literacy test for engineering students. *Journal Turkish Sci Educ*, *13*(3), 185–198.

Treagust, D. F., G, T., & Chia. (2002). Development and Application of a Two-Tier Multiple Choice Diagnostic Instrumen to Assess High School Student's Understanding of Inorganic Chemistry Qualitative Analysis. *Journal of Reaserch in Science Teaching.*, *39*(4), 283–301.

Treagust, D., & Haslam, F. (1986). Evaluating Secondary Students Misconceptions of Photosyntesis and Respiration in Plants Using A two-tier Diagnostik Instrumen. *Educational Resources Information Center (ERIC)*, 5.

Usmaedi. (2021). "Education Curriculum For Society 5.0 In The Next Decade,." *jurnal pendidik dasar setiabudhi*, *4*(2), 63–79.

Yuliaristiawan, E. D. (2023). *Mengenal Tes Diagnostik dalam Pembelajaran*. https://smpbss.sch.id/2023/09/17/mengenal-tes-diagnostik-dalam-pembelajaran/