EduChemia
Jurnal Kimia dan Pendidikan

Vol.10, No.1, 2025, p. 1-13
e-ISSN 2502-4787

AI-POWERED AUTOMATED ASSESSMENT: AIKEN INDEX ANALYSIS OF CONTENT VALIDITY IN SCIENTIFIC LITERACY INSTRUMENTS

Bela Anisa Putri[1], Sri Yamtinah[2], Ari Syahidul Shidiq[3], Hayuni Retno Widarti[4], Antuni Wiyarsi[5]

[1,2,3]*Pendidikan Kimia, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Sebelas Maret, Surakarta, Indonesia*
[4]*Pendidikan Kimia, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Negeri Malang, Malang, Indonesia*
[5]*Pendidikan Kimia, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia*

E-mail: *putribelaanisa@gmail.com; **jengtinah@staff.uns.ac.id; ***arishidiq@staff.uns.ac.id; ****hayuni.retno.fmipa@um.ac.id; *****antuni_w@uny.ac.id

Received: 01 Januari 2025. Accepted: 19 Februari 2025. Published: 31 Maret 2025

DOI: 10.30870/educhemia.v10i1.30640

Abstract: This study aims to analyze the content validity of the Scientific Literacy Asesment (SLA) instrument, which integrates ethnochemistry based on artificial intelligence through expert agreement based on the Aiken index so that the instrument can measure what should be measured. The research method used is a qualitative descriptive method based on the results of content validity calculated through the Aiken formula. Content validity data were obtained from 10 experts, namely, chemistry education lecturers from UNS, UNY, and UM and high school chemistry teachers in Surakarta, through focus group discussions (FGDs). Content validity is assessed by a score of 1 – 4, namely, irrelevant (TR) with a score of 1; less relevant (KR) with a score of 2; quite relevant (CR) with a score of 3; and relevant (R) with a score of 4 for each question item with aspects measured, namely, aspects of content, language, and construct. Based on the calculation, 15 questions were declared valid with an Aiken index value ≥ 0.73.

Keywords: Aiken, SLA, stoichiometry, content validity

Abstrak: Penelitian ini bertujuan untuk menganalisis validitas isi instrumen Scientific Literacy Asesment (SLA) terintegrasi etnokimia berbasis Artificial Intelligence melalui kesepakatan para ahli berdasarkan indeks Aiken, sehingga instrumen tersebut dapat mengukur apa yang seharusnya diukur. Metode penelitian yang digunakan adalah metode deskriptif kualitatif berdasarkan hasil validitas isi yang dihitung melalui formula Aiken. Data validitas isi diperoleh dari 10 orang ahli, yaitu dosen pendidikan kimia dari UNS, UNY, dan UM serta guru kimia SMA di Surakarta melalui focus group discussion (FGD). Validitas isi dilakukan dengan cara memberikan skor 1 - 4, yaitu tidak relevan (TR) dengan skor 1, kurang relevan (KR) dengan skor 2, cukup relevan (CR) dengan skor 3, dan relevan (R) dengan skor 4 pada setiap butir soal dengan aspek yang diukur yaitu aspek isi, bahasa, konstruk. Berdasarkan hasil perhitungan, 15 butir pertanyaan dinyatakan valid dengan nilai indeks Aiken ≥ 0,73.

Kata kunci: Aiken, SLA, stoikiometri, validitas isi

INTRODUCTION

One of the crucial needs of students in the 21st century is science literacy (Triwibowo et al., 2021). According to (Zulfa & Haryanto, 2021), science literacy aims to form individuals who are accustomed to thinking critically, are sensitive to the environment, and can apply scientific knowledge to solve real problems faced by many people. Based on the results of the Program for International Student Assessment (PISA) survey 2022 (PISA, 2022), Indonesia experienced a decline in science literacy scores from 396 in 2018 to 383, placing it in the 15th lowest rank out of 81 countries. This science literacy includes literacy in chemistry learning (Rahmadani et al., 2018).

Based on the results of interviews with teachers in Surakarta, in chemistry learning, teachers have applied ethnochemistry in explaining the material to students. The material taught is often closely related to culture and daily life. Ethnochemistry is a branch of ethnoscience that connects chemistry with local traditions to make learning more interesting for students. By studying chemistry in a cultural context, students not only learn theory but also see its application in everyday life to create a more meaningful learning experience (Aldiansyah et al., 2023). However, the questions presented as assessment instruments do not involve narratives related to ethnochemistry, so students lack relevant context.

After learning is complete, teachers always conduct assessments via assessment instruments that aim only to measure students' concept understanding according to the desired learning achievement criteria. In the assessment instrument, the teacher does not involve the measurement of science literacy skills in the assessment. Therefore, an assessment instrument that is not only able to measure students' understanding of the material but also able to measure students' science literacy skills related to ethnochemistry in the context of local culture is needed. This is supported by the opinion of Toharudin (Windyariani et al., 2021), who stated that the understanding of science learning aimed at shaping science literacy in students is still not well understood by science teaching teachers. The learning process and evaluation tools used tend to be conventional and focus on mastering concepts, so students are not familiar with science literacy skills.

Chemistry is complex and abstract, causing many students to consider chemistry a difficult subject (Yamtinah et al., 2022). Stoichiometry is a chemical material that requires a high degree of reasoning (Supatmi, 2022). Based on the results of interviews with teachers in Surakarta, stoichiometry questions rarely involve narration because the questions are presented only directly in the form of orders. This causes a lack of context that can help students understand the application of concepts in real situations. Therefore, questions with narratives are more effective because they can train students to read deeply to understand questions that are conveyed implicitly through narratives. In addition, with the narrative, students gain new knowledge related to the application of stoichiometry in real life.

Students' science literacy skills were measured using the scientific literacy assessment (SLA), which is designed with science literacy indicators. SLA is a tool designed to evaluate science literacy to measure demonstrated knowledge and motivational beliefs about science learning (Stylos et al., 2023). According to (Shwartz et al., 2006), science literacy in chemistry consists of 4 aspects, namely, content aspects, context aspects, higher-order learning skills (HOLS) aspects, and attitudinal aspects. The development of SLA tools often involves a systematic theoretical review to conceptualize science literacy (Istyadji & Sauqina, 2023).

SLA instruments can be realized in the form of multiple-choice questions, descriptions, or other forms of questions (Islamiyah et al., 2024). A description test is also called an essay test or a subjective test. The description test requires test takers to compile their own answers (Widoyoko, 2012). Descriptive tests are very suitable for assessing the results of complex learning processes, which are difficult to measure with objective tests (Putri et al., 2022). The descriptive test allows teachers to observe and analyze the ability of students in logic so that they can determine the ability of science literacy from the composition of answers provided by students (Arikunto, 2013).

The results of the interviews with high school teachers in Surakarta revealed that the description questions had a specific assessment rubric, which caused teachers to take a long time to assess them. This is also supported by (Ismail & Vita 2020), who stated that one of the weaknesses of description questions is that correcting essay tests takes quite a long time and drains more energy because each answer must be read carefully one by one. However, artificial intelligence can

analyze results in real time and provide direct feedback to students, including grades and analyses of answer errors. AI is an intelligent system created by humans and integrated into machines, allowing the machine to think and make decisions (Azizah et al., 2021). Therefore, it is necessary to integrate AI in the development of SLA instruments.

In the development of test instruments, validity is an important consideration. Validity describes the extent to which the measuring instrument is able to measure what should be measured (Azwar, 2000). Validity is grouped into 3 types, namely, criterion validity, content validity, and construct validity. Content validity is carried out by experts who aim to assess the suitability of indicators with the purpose of instrument development, see the suitability of questions with question indicators, see the correctness of the concept of question items, see the correctness of the answer key, and the correctness of language (Retnawati, 2016). Content validity refers to the extent to which the test content of the characteristics to be measured (Yamtinah et al., 2016). Content validity was quantified through the validation assessment sheet, which was then used to calculate the expert agreement index using the Aiken or Gregory index.

In this study, the SLA instrument developed was integrated with ethnochemistry to create a relevant assessment instrument in accordance with the existing methods. The advantage of the ethnochemically integrated SLA instrument according to the characteristics of ethnochemistry is that the question narrative is able to link chemical theory with the local cultural context so that it can enrich students' understanding and connect science with local culture (Aldiansyah et al., 2023). The use of AI also helps in the development of sophisticated and effective evaluation systems. In this SLA instrument, the AI model used is Gemini, which has been proven to be able to improve the system's ability to assess applied mathematics and provide meaningful and context-specific feedback (Campino, 2024). The SLA instrument that has been developed needs to be tested for validity.

Therefore, this study measured the content validity of the AI-based ethnochemically integrated SLA instrument with Aiken's index analysis through an assessment of a number of instrument items by 10 validators with a total of 3 assessment aspects, namely, content, language, and construct aspects. Based on the number of validators and assessment aspects, the instrument is said

to be valid if Aiken's index has exceeded the minimum limit set by (Aiken, 1985), namely, 0.73. In the implementation of content validity, validators are given a package of instruments that include question indicators, science literacy items, and assessment sheets.

METHOD

This type of research is research and development (R&D), which is a process for developing and validating educational products, as described by Borg and Gall (1983). R&D involves several stages, namely, studying relevant research findings, designing the product, testing the product in its use environment, and revising it (Borg and Gall, 1983). There are three models of development research: procedural, conceptual, and theoretical. The descriptive procedural model shows the steps of making the product, the analytical conceptual model analyzes the product components, and the theoretical model describes the framework of thinking based on theory and empirical data (Punaji, 2010). This research uses a procedural model to produce a product that has been tested for feasibility through a content validity process on the product developed, namely, the scientific literacy assessment (SLA) instrument integrated with ethnochemistry based on artificial intelligence (AI) on stoichiometric materials for class XI SMAs/MAs.

The method used in this research is descriptive qualitative, which is based on the results of content validity calculated through the Aiken formula. In this study, content validity was assessed by ten experts, including lecturers from Universitas Sebelas Maret (UNS), Universitas Negeri Malang (UM), Universitas Negeri Yogyakarta (UNY), and high school chemistry teachers in Solo city. The content validity process was conducted through a focus group discussion (FGD) held at Sebelas Maret University. The experts provided an assessment by giving a score of 1 – 4 for each question item based on three aspects measured, namely, content, language, and construct aspects.

The content aspect refers to the extent to which the material in the questions is relevant to the topic being taught, including the accuracy of the information on ethnochemistry that has been integrated. The questions should reflect the content of the lessons learned by the students, and the narrative should not lead to wrong perceptions. The language aspect relates to the clarity, suitability and accuracy of the language used in accordance with Indonesian language rules. Questions should be

written in a language that is easily understood by students without ambiguity. The construct aspect focuses on how the instructions in the questions are conveyed clearly and easily understood so as not to cause various interpretations.

The content validity of the instrument is determined on the basis of the agreement of experts, who are believed to be able to determine whether the instrument is accurate in measuring the intended ability. This agreement is used to ensure that the test or nontest instrument is valid and reliable. Aiken's validity index is used to measure the level of agreement among experts, which ensures that the instrument developed can actually measure the intended purpose. Aiken's formula (Aiken, 1985) is used to calculate the level of validity based on expert judgment.

$$V= \Sigma S / [n(C-1)]$$

with S= R-Lo

Description:

V: Aiken's validity index

S: the score given by the rater minus the lowest score in the category

R: score given by the rater

Lo: lowest assessment score (1)

C: highest assessment score (4)

n: number of validators (assessors)

Through this formula, the researcher can calculate the level of agreement among experts regarding the relevance of each item. This process helps determine the overall content validity of the developed instrument. After the level of agreement was calculated, the obtained validity index was compared with the table V value proposed by Aiken (1985) to determine whether the instrument was valid. This table V value varies depending on the number of answer categories used in the assessment and the number of validators or raters involved in the validation process.

RESULTS AND DISCUSSION

According to (Retnawati, 2016), in the development of test instruments, several steps need to be taken. The first stage is to determine the purpose of preparing the instrument, followed by finding relevant theories to underlie the test. Furthermore, the next step is to compile indicators of questions and then design items that are in accordance with these indicators. After that, the validity stage is carried out to ensure that the instrument measures what should be measured, followed by the revision stage to improve the instrument based on the validity results. The last stage involves testing the respondents, analyzing the reliability, determining the level of difficulty, analyzing the differentiating power, and finally assembling the test

instrument ready for use. Validity is one of the important steps in test development.

Validity can be divided into three types: criterion validity, content validity, and construct validity. Content validity is carried out by experts who aim to assess the suitability of indicators with the purpose of instrument development, see the suitability of questions with question indicators, see the correctness of the concept of question items, see the correctness of the answer key, and the correctness of language (Retnawati, 2016). Content validity ensures that instrument items can accurately describe the abilities that should be measured (Neuman, 1994). Thus, the content validity of an instrument relates to the extent to which the items can cover the abilities that should be measured.

In this study, content validity was assessed by 10 assessors, with 15 description questions assessed from the aspects of content, language, and construct. The following indicators for each aspect assessed are presented in Table 1.

Table 1. Indicator of validation assessment aspects

Aspect	Indicator
Content	1. Relevance of the questions to the material taught 2. The suitability of the concepts and information presented in the question items with the knowledge possessed by students 3. The correctness of the facts in the question items
Language	1. Grammatical accuracy in question items 2. Clarity of sentence structure to facilitate student understanding 3. Accuracy of spelling, punctuation, and word usage in question items 4.
Construct	1. Clarity of instructions 2. Clarity of the image presented in the problem 3. Use of visual aids or additional information provided to assist student understanding

The aspect indicators in Table 1 clarify that the assessment evaluated the suitability of the integrated instrument, which was developed through checking the relevance of the questions to the ethnochemical material, the suitability of the information presented to the students' knowledge, and the truth of the facts related to local culture and chemical concepts. In addition, the accuracy of grammar, clarity of sentence structure, and use of proper spelling ensure students' understanding of the relationship between chemical concepts and culture. The question instructions, pictures, and visual aids or additional information also serve to support students' understanding of ethnochemistry so that the whole question can be used to assess the integration between chemistry and traditional knowledge effectively.

Qualitative data obtained from the instrument validation results were then analyzed to evaluate the assessment and views of the AI-based ethnochemistry-integrated SLA instrument products on stoichiometric materials that had been made. The analysis is based on information that has been obtained through validation results containing suggestions, comments, and input from validators to improve the instruments that have been developed. From the FGD activities that have been carried out, the content validity values of 10 assessors are obtained via the Aiken formula, which can be seen in Table 2.

Table 2. Results of Aiken Index Analysis of the AI-based ethnochemistry-integrated SLA Instrument

Question	V Values	V Table	Conclusion
1	0,84	0,73	Valid
2	0,84	0,73	Valid
3	0,88	0,73	Valid
4	0,90	0,73	Valid
5	0,90	0,73	Valid
6	0,99	0,73	Valid
7	0,94	0,73	Valid
8	0,91	0,73	Valid
9	0,89	0,73	Valid
10	0,96	0,73	Valid
11	0,98	0,73	Valid
12	0,94	0,73	Valid
13	0,91	0,73	Valid
14	0,90	0,73	Valid
15	0,96	0,73	Valid

Content validity relates to how well the fit between question items and indicators of science literacy skills is. In this study, four answer categories were used to determine validity, namely, irrelevant (TR), with a score of 1; less relevant (KR), with a score of 2; moderately relevant (CR), with a score of 3; and relevant (R), with a score of 4. In this study, content validity is based on Aiken's formula (Aiken, 1985) through the Aiken index value. According to Aiken (1985), the validity of a question item is considered good if the Aiken index (V index) is greater than or equal to 0.73 based on the assessment of ten raters who give scores in four answer categories.

The Aiken index value (V index) shows the suitability of the item with the indicator to be measured using the item based on the rater agreement index (Retnawati, 2016). Based on the calculation results, 15 items from the SLA instrument were declared valid. Table 1 displays the results of the calculation of Aiken's index on the SLA instrument with 15 items of high validity. Aiken's validity index is declared low if it is 0.4, medium if it is 0.4 – 0.8, and high if it is above 0.8 (Aiken, 1985). The question item is said to improve if the Aiken index value obtained is close to a score of 1.0, which indicates that the question item is increasingly relevant to the predetermined indicators (Retnawati, 2016).

The validation results from 10 validators produced several suggestions

and comments, which are summarized below. 1) The language used in questions and narratives should be easy to understand, not lead to multiple interpretations, and not cause misconceptions among students. Therefore, the author should ensure that the questions and narratives are presented in a clear and unambiguous manner and involve competent parties to check for potential misinterpretations. 2) Editing in writing narratives and questions must be in accordance with the General Guidelines for Indonesian Spelling (PUEBI). This is important so that the text does not cause confusion owing to grammatical or spelling errors. 3) The material presented in the narrative should be simplified to make it easier for students to understand. The use of simpler language and the addition of concrete examples can help students understand the content of the narrative. 4) The questions should be closely related to the narrative so that students can answer the questions based on the information they have read. 5) Question indicators must also be relevant to the learning objectives so that each question truly measures the competencies to be achieved. 6) The numbers used in the questions must be simple enough and easy to calculate without the help of a calculator to match the cognitive abilities of students at a certain level of education. 7) Pictures or illustrations that support the narrative should be included, as this can help students visualize the material provided. The images should be relevant and support the understanding of the material, not just decorative. 8) The narrative, questions, and answer key should use consistent units. Inconsistency in the use of units can cause confusion and reduce accuracy in answering questions. 9) In terms of scoring, the number of scores on each question should be uniform to make it easier for the AI system to analyze, as well as to ensure equality in the weight of each question.

Based on the suggestions and comments from the experts, 15 question items on the SLA instrument have been declared valid but still require improvement to improve their quality. Therefore, it is necessary to revise each question item through the suggestions and input provided by the validators. The revised SLA instrument is then applied through an AI-based system. Therefore, it is necessary to have a supporting prompt to obtain the expected results. The expected result is that students receive a total score with details of the score obtained accompanied by an explanation of why they receive the score and receive feedback to evaluate the answers that have

been obtained. The following is the prompt format that has been developed in Table 3.

Table 3. Criteria of Student Understanding Level

Sub Content	Information
Role and Objective	Contains an overview of what role AI should play in order to evaluate students' answers.
Question	Contains questions
Answer Key	Contains answer key
Scoring Rubric	Contains detailed scoring guidelines. In this system, the maximum score for each question has the same value to make it easier for the AI system to work.
Evaluation Instructions	Contains details regarding how the AI will provide evaluation of student answers.
Example Feedback Structure	Contains examples of feedback structures that students receive. Here is the format: Total Score: [Insert Total Score Here] out of 10 points Scoring Breakdown: Question : [Insert Score] - [Reason for Score] Feedback: "..."
Final Comment	Contains a conclusion of how the student performed along with highlighting strengths and areas that need improvement.

CONCLUSION

This study aims to analyze the content validity of the scientific literacy assessment (SLA) instrument integrated with ethnochemistry based on expert agreement using the Aiken index. Content validity data were obtained from 10 experts, including chemistry education lecturers and high school chemistry teachers, through focus group discussions (FGDs). The experts rated each question item on aspects of content, language, and construct. Based on the calculation, 15 questions were declared valid with an Aiken index value ≥ 0.73.

ACKNOWLEDGMENT

The authors express their gratitude to the Institute for Research and Community Service (LPPM) of Universitas Sebelas Maret for supporting the Collaborative Research program between Universitas Sebelas Maret (UNS), Universitas Negeri Malang (UM), and Universitas Negeri Yogyakarta (UNY) under the Indonesia Collaboration Research (RKI) scheme with contract number 285/UN27.22/PT.01.03/2024. This support has significantly contributed to the successful implementation of this research.

REFERENCES

Aiken, L. R. (1985). Three Coefficients For Analyzing The Reliability And Validity Of Ratings, Educational And Psychological Measurument.

Educational and Psychological Measurement, *45*(1), 131–142. https://doi.org/10.1177/0013164485451012

Aldiansyah, A., Pasa, J. I., Muttaqin, M. R., Awaliyah, N. N., & Erika, F. (2023). Literatur Review : Keterkaitan Pembelajaran Kimia Terhadap Pendekatan Etnokimia Di Indonesia. *CHEDS: Journal of Chemistry, Education, and Science*, *7*(2), 238–246. https://doi.org/10.30743/cheds.v7i2.8416

Arikunto, S. (2013). Contribution Of Educational Management To Optimal Service Of Educational Development. *PSIKOPEDAGOGIA Jurnal Bimbingan Dan Konseling*, *2*(2), 70. https://doi.org/10.12928/psikopedagogia.v2i2.2573

Azizah, D., Wibawa, A., & Budiarto, L. (2021). Hakikat Epistemologi Artificial Intelligence. *Jurnal Inovasi Teknologi Dan Edukasi Teknik*, *1*(8), 592–598. https://doi.org/10.17977/um068v1i82021p592-598

Azwar, S. (2000). *Reliabilitas Dan Validitas*. Pustaka Pelajar Offset.

Borg, W And Gall, M. (1983). *Educational Research*. Longman.

Campino, J. (2024). Unleashing The Transformers: NLP Models Detect AI Writing In Education. *Journal of Computers in Education*, *0123456789*. https://doi.org/10.1007/s40692-024-00325-y

Islamiyah, E. F. N. I., Dewanti, B. A., & Santoso, A. (2024). Development Of Science Literacy Assessment Instruments In Science Learning Using Contextual Learning Models For Junior High School. *Quantum: Jurnal Inovasi Pendidikan Sains*, *15*(1), 75–89. https://doi.org/10.20527/quantum.v15i1.17255

Ismail, M. I., & Vita, P. (2020). *Evaluasi Pembelajaran: Konsep Dasar, Prinsip, Teknik dan Prosedur*. Rajawali Pers.

Istyadji, M., & Sauqina. (2023). Conception Of Scientific Literacy In The Development Of Scientific Literacy Assessment Tools: A Systematic Theoretical Review. *Journal of Turkish Science Education*, *20*(2), 281–308. https://doi.org/10.36681/tused.2023.016

Neuman, W. L. (1994). *Social Research Methods Qualitative And*

Quantitative Approaches. (2nd ed.). Allyn And Bacon.

PISA. (2022). Notas por país: México. (2024). *Perfiles Educativos*, *46*(183), 188–202. https://doi.org/10.22201/iisue.24486167e.2024.183.61714

Punaji, S. (2010). *Metode Penelitian Pendidikan dan Pengembangan.* Kencana.

Putri, H., Susiani, D., Wandani, N. S., & Putri, F. A. (2022). Instrumen Penilaian Hasil Pembelajaran Kognitif Pada Tes Uraian Dan Tes Objektif. *Jurnal Pendidikan Dan Ilmu Sosial (Jupendis)*, *2*(4), 139–148. https://doi.org/10.54066/jupendis.v2i4.2159

Rahmadani, Y., Fitakurahmah, N., Fungky, N., Prihatin, R., Majid, Q., & Prayitno, B. A. (2018). Profil Keterampilan Literasi Sains Siswa Di Salah Satu Sekolah Swasta Di Karanganyar. *Jurnal Pendidikan Biologi*, *7*(3), 183. https://doi.org/10.24114/jpb.v7i3.10123

Retnawati, H. (2016). Proving Content Validity Of Self-Regulated Learning Scale (The Comparison Of Aiken Index And Expanded Gregory Index). *Research and Evaluation in Education*, *2*(2), 155–164. http://dx.doi.org/10.21831/reid.v2i2.11029

Shwartz, Y., Ben-Zvi, R., & Hofstein, A. (2006). The Use Of Scientific Literacy Taxonomy For Assessing The Development Of Chemical Literacy Among High-School Students. *Chemistry Education Research and Practice*, *7*(4), 203–225. https://doi.org/10.1039/B6RP90011A

Stylos, G., Siarka, O., & Kotsis, K. T. (2023). Assessing Greek Pre-Service Primary Teachers' Scientific Literacy. *European Journal of Science and Mathematics Education*, *11*(2), 271–282. https://doi.org/10.30935/scimath/12637

Supatmi, S. (2022). Peningkatan Keterampilan Proses Sains Melalui Praktikum Kimia Berbasis Skala Mikro Materi Stoikiometri. *Jurnal Riset Pendidikan Kimia*, *12*(1), 47–58. https://doi.org/10.21009/JRPK.121.07

Triwibowo, S., Parno, P., & Latifah, E. (2021). Kemampuan Literasi Sains

Siswa SMA Topik Usaha Dan Energi. *Jurnal Pendidikan: Teori, Penelitian, Dan Pengembangan*, *6*(11), 1787. https://doi.org/10.17977/jptpp.v6i11.15146

Widoyoko, E. P. (2012). *Teknik Penyusunan Instrumen Penelitian*. Pustaka Pelajar.

Windyariani, S., Setiono, & Sutisnawati, A. (2021). Pengembangan Model Asesmen Literasi Sains Berbasis Konteks Bagi Siswa Sekolah Dasar. *Seminar Nasional Kedua Pendidikan Berkemajuan Dan Menggembirakan (The Second Progressive and Fun Education Seminar)*, 279–286.

Yamtinah, S., Dewi, M. C., Nurhayati, N. D., Susilowati, E., Fakhrudin, I. A., Ramadhani, D. G., Shidiq, A. S., & Saputro, S. (2022). Content Validity In Android-Based Augmented Reality Media For High School Science Students On Covalent Bonds Topic: Rasch Model Analysis. *Jurnal Pendidikan Sains Indonesia*, *10*(2), 240–249. https://doi.org/10.24815/jpsi.v10i2.23280

Yamtinah, S., Saputro, S., & Utami, B. (2016). Content Validity And Scoring Of Two Tier As Measuring Instrument Of Science Process Skills For Knowledge Aspects In Chemistry Learning. *Prosiding ICTTE FKIP UNS*, *2*, 911–916.

Zulfa, L. N., & Haryanto, H. (2021). Pengaruh Media Macromedia Flash Terhadap Literasi Sains Dan Sikap Demokratis Mahasiswa. *Jurnal Pendidikan Sains Indonesia*, *9*(1), 52–64. https://doi.org/10.24815/jpsi.v9i1.18266