Jurnal Pembelajaran Dan Biologi Nukleus p-ISSN: 2442-9481
Vol 11 (2): 769 - 784, June 2025 e-ISSN: 2685-7332

Development of Science Technology Engineering and Mathematics (STEM) Embedded Evaluation Tools for Micology Learning

Nurul Alfi Oetari, Idramsa Idramsa(*)

Biology Education Study Program, Mathematics and Natural Sciences Faculty,
State University of Medan
Jl. William Iskandar Ps. V, Kenangan Baru, Percut Sei Tuan, Deli Serdang,
North Sumatera 20221, Indonesia

***Corresponding Author:** idramsa@unimed.ac.id

Submitted March 24 Th 2025 and Accepted May 30 Th 2025

Abstract

Background: *STEM learning can help students understand science concepts, improve science skills, create active, creative, motivating learning and provide opportunities for students to apply the material and build their knowledge. This research aims to develop valid and reliable science, technology, engineering and mathematics-based student evaluation tools on fungi material.* **Methodology:** *The evaluation instrument was developed using the 4D development model, which includes four stage: Define, Design, Develop and Determine. The study involved 30 first-grade students from MAS Madinatussalam, North Sumatra. The developed evaluation tool consisted of 40 questions. Validation by industry experts confirmed the instrument's validity and reliability.* **Findings:** *Expert evaluations of the questions indicated high suitability, with an average score of 96.75%. Material experts rated the instrument with a suitability score of 92.5%, while media experts gave an average score of 86.1%, all falling within the "very good" category. Reliability testing using Cronbach's Alpha yielded a score of 0.886, indicating high reliability. Based on these results, the STEM-based evaluation tool is considered both valid and reliable for classroom use.* **Contribution:** *Expected to provide information as a basis for consideration, supporting the improvement of problem-solving skills, increasing critical and creative thinking skills.*

Keywords: *Reliability; Student Evaluation; Stem; Validity*

https://doi.org/10.36987/jpbn.v11i2.7235

INTRODUCTION

Implementing assessment in the 2013 curriculum directs teachers to conduct holistic assessments that cover three domains of ability: affective, psychomotor, and cognitive. The cognitive domain is related to knowledge or understanding, the affective domain is related to attitudes and interests, and the psychomotor domain is related to skills. Assessment in the 2013 curriculum is intended so that students do not feel pressured in participating in learning (Andriani, 2021). Assessment of learner learning outcomes includes the competencies of attitude, knowledge and skills which are carried out in a balanced manner so that they can be used to determine the relative position of each learner to predetermined standards (Majid, 2015). The curriculum refers to a structured plan that outlines learning objectives, subject matter, instructional materials, and teaching methods, serving as a guide for organizing educational activities aimed at achieving specific learning outcomes (Mendiknas, 2003).

Assessment is intended to determine the extent to which students are able to achieve learning objectives. The process of collecting data in an assessment can be helpful for understanding the acquisition of student learning outcomes. Minister of Education and Culture Regulation Number 66 of 2013 on Educational Assessment Standards encompasses several types of assessments, such as authentic assessment, self-assessment, portfolio assessment, daily tests, end-of-semester examinations, competency level tests, national examinations, and school-based exams (Mawardi, 2021).

The 2019 results of the Program for International Student Assessment (PISA) indicated that Indonesia ranked among the bottom six out of 77 countries regarding student learning outcomes (PISA, 2019). In December 2023, PISA released updated data showing that Indonesia ranked 12th lowest in numeracy, 11th lowest in literacy, and 15th lowest in science, out of 81 countries. In all three areas, Indonesian student's performance was below the average score of all participating countries (PISA, 2023). Compared to Singapore, Malaysia, Brunei Darussalam and Australia, Indonesia ranked much lower. This gap in performance suggests that need for improvement (Suparya et al, 2022). To tackle the issues of low literacy rates and educational disparities, the Indonesian government has made the literacy movement a key priority through the Ministry of Education and Culture. Scientific literacy is one of the essential skills that students need to master, as it is crucial for preparing them to face future challenges (Idawati, 2019).

One approach that can be used for the learning process is STEM (Science, Technology, Engineering and Mathematics). The STEM approach combines four branches of science components into one, namely science, technology, engineering and mathematics. The STEM approach is suggested to help the success of 21st century skills. In this case, it means that through the STEM approach, students not only memorise concepts, but also understand and comprehend science concepts related to everyday life (Simatupang, 2019).

If STEM learning is appropriate, it will affect the success of students in understanding the material and can also improve skills in the science process. The STEM learning model can also create an active, creative learning and can motivate

students to find a concept in learning and provide opportunities for students to apply the material, build their knowledge and work in groups so as to develop scientific attitudes and skills possessed by students (Herak & Lamanepa, 2019). Kristanto et al., (2021) revealed that the STEM-based learning outcomes test instrument developed was effective, because it could measure students' science process skills in learning outcomes evaluation activities and received positive responses from students.

An evaluation tool is something that is used to make it easier for someone to perform a task or achieve a goal more effectively and efficiently. An evaluation tool is called good if it is able to evaluate something with results like the situation being evaluated (Arikunto, 2012). Evaluation is also an important part of the learning process because it is one of the benchmarks of learning success. Evaluation is one of the illustrations of the success of the learning process carried out by educators (Purnamasari & Rochmawati, 2015). valuation is useful as a reference for making a decision or policy (Febriana, 2019).

A test is a tool or method applied to conduct measurement activities, consisting of various questions or a set of tasks that students must complete or respond to, aiming to evaluate specific aspects of their behavior (Arifin, 2016). It is primarily used to assess how well a learner has understood and mastered the instructional content, particularly in terms of knowledge and skills (Sudaryono, 2012). Within certain parameters, tests can also serve to evaluate learning outcomes in the affective and psychomotor domains (Sudjana, 2016).

A brief overview of micology learning that Fungi cannot photosynthesize and live saprophytically, and their breeding tools are spores and hyphae. The part of the fungus that is seen appearing above the surface of the soil is only the fruit of the fungus. the other part of the fungus is embedded in the soil in the form of thick, chaotic white threads. These threads are called mycelium (Safitri, 2020). Given that micology represent a complex biological concept requiring not only cognitive understanding but also process skills to observe and analyze, this topic is considered highly relevant for developing a STEM-based assessment instrument. Therefore, the present study seeks to construct and validate such an instrument to comprehensively measure students' cognitive and scientific process skills in accordance with the 2013 curriculum.

METHODS

The development approach applied in this research is the 4D model (Define, Design, Development and disseminate) introduced by Thiagarajan et al., (1974). The research method used is the Research and Development (R&D) method, which is intended to produce specific products and assess their effectiveness.

Instrument

Research instruments are a vital component of any study, as the main objective of research is to gather data. The instruments utilized in this study consist

of: (1) a Test Instrument Validation Sheet; (2) a Teacher Response Questionnaire; and (3) a Student Response Questionnaire.

Data Collection

The data collection techniques used are observation, validation questionnaire, teacher response questionnaire, learner response questionnaire and multiple choice tests. The questionnaires used in this study are validation questionnaires by expert lecturers on evaluation questions, validation questionnaires by material expert lecturers, validation by media expert lecturers, validation questionnaires by biology study teachers, teacher response questionnaires and student response questionnaires.

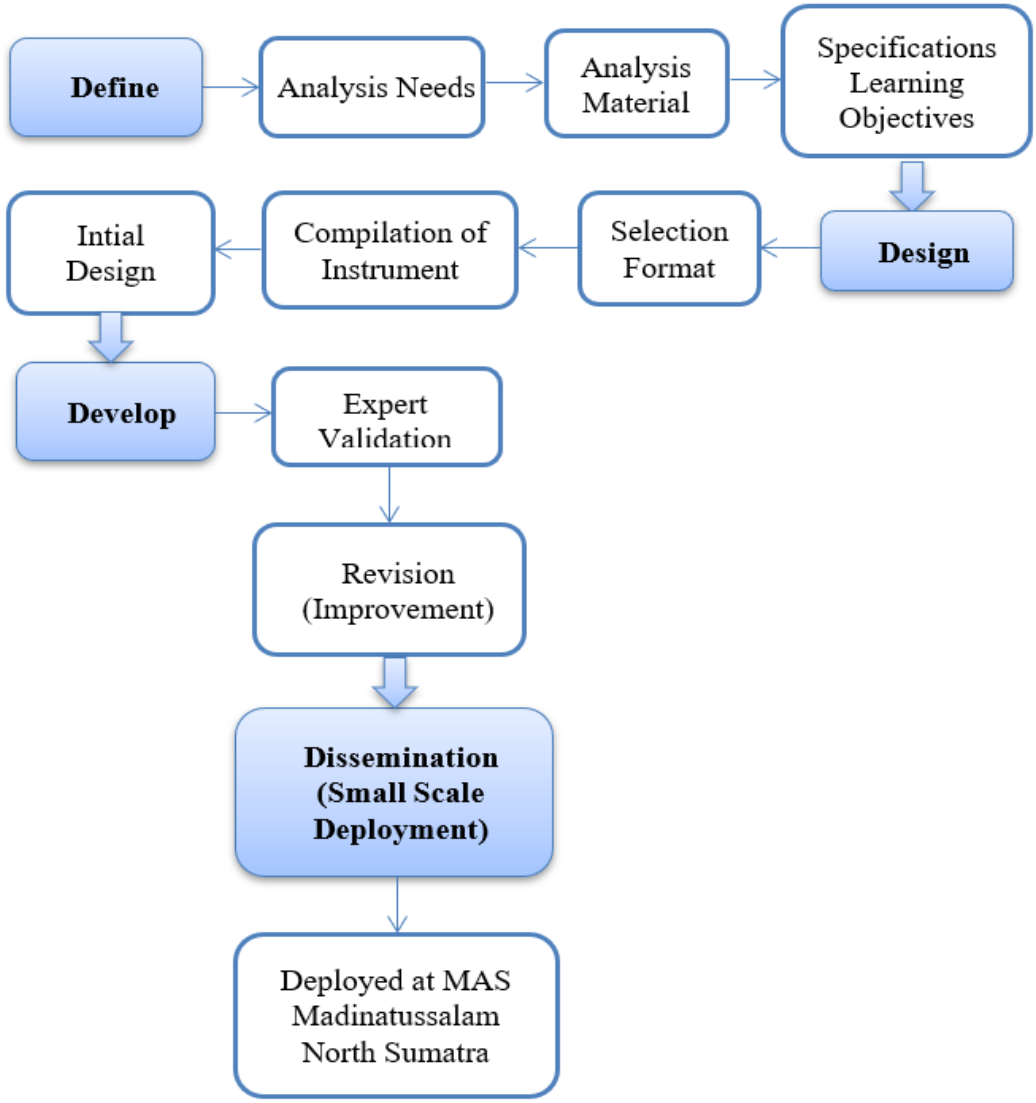

Figure 1. 4D Development (Thiagarajan et al., 1974)

Procedure

The definition stage is the initial stage whose purpose is to determine and define the needs in the learning process as well as collect various information related to the product to be developed. In the defining stage, needs analysis, material analysis, and specification of learning objectives are carried out. Conducting interviews with biology teachers and students. Aiming to find out the extent of the evaluation of biology subjects on micology learning (fungi) in first grade-X class of MAS Madinatussalam North Sumatra. The design stage aims to obtain an initial design of the instrument developed, carried out by determining the form of the instrument totaling 40 items in the form of multiple choice questions, with answer choices from a, b, c, d, and e, with the cognitive domain C4 - C6. The development phase involves creating the student evaluation tool based on the initial design, followed by validation from content experts, assessment specialists, and media professionals. In this development stage, the aim is to obtain a good test instrument and have good eligibility standards as well. This dissemination stage is the stage of disseminating products that are suitable for use. After proving to be ideal for use, the product is ready for broader use. At this stage of dissemination has the aim that the validated test of the learning device is tested and improvements have been made, after which it is distributed to the field (Kurniawan, 2017).

Data Analysis

The techniques used to analyze the data in this study are instrument validity analysis, instrument reliability analysis, difficulty level analysis, differentiator analysis, and teacher and learner response analysis. The learning outcomes test that has been validated by experts who will later calculate the results then look for the average score of the experts' assessment. The experts' assessment sheet is then analyzed to determine the quality of the product the researcher has made. The following formula for calculating the validity of evaluation tools based on expert judgment refers to Febriani et al., (2021) :

$$Vah = \frac{Tse}{Tsh} \text{ X } 100\ \% \quad \text{..... (1)}$$

Description:
Vah : Expert Validation
Tse : Total score to be achieved
Tsh : Total expected score

Table 1. Expert Validation Sheet Criteria, Refers to Febriani et al., (2021)

Value Achievement Criteria (%)	Classification
≥ 81,5-100	Strongly Agree
≥ 62,5-81,5	Agree
≥ 43,5-62,5	Moderately
≥ 25-43,5	Agree Disagree
≥ 0-25	Strongly Disagree

Table 2. Expert Validation Sheet Score, Refers to Febriani et al., (2021)

Score	Answer Options
5	Strongly Agree
4	Agree
3	Moderately
2	Agree Disagree
1	Strongly Disagree

Table 3. Interpretation of Reliability Values, Refers to Arikunto (2017)

Value r	Interpretation
0,80-1,00	Very High
0,60-0,79	High
0,40-0,59	Simply
0,20-0,39	Low
<0,200	Invalid

Tabel 4. Distinguishing Power Index Criteria, Refers to Arikunto (2017)

Distinguishing Power Index	Criteria
0,40 - 1,00	Good
0,30 - 0,39	Accepted but needs improvement
0,20 - 0,29	Fixed
0,19 - 0,00	Not used/discarded

Tabel 5. Criteria for Question Item Difficulty Index, Refers to Arikunto (2017)

Difficulty Level	Value Criteria
Difficult	P < 0,3
Medium	0,3 ≤ p ≤ 0,7
Easy	P > 0,7

In analyzing the responses of teachers and students, researchers measured by giving a response questionnaire to the teacher and each student using a Likert scale. With the following formula referring to Febriani et al., (2021) :

$$P = \frac{f}{N} \text{X } 100\ \% \quad \text{..... (2)}$$

Description:
P = Presentase Rate
F = Raw Score Obtained
N = Maximum Score

Table 6. Teacher and Learner Response Score Refers to Febriani et al., (2021)

Score	Answer Options
5	Strongly Agree
4	Agree
3	Moderatel
2	Agree Disagree
1	Strongly Disagree

RESULTS AND DISCUSSION

In this study, the product developed is a STEM (Science, Technology, Engineering and Mathematics) based student evaluation tool on fungi material. The instrument amounted to 40 multiple choice questions and based on the cognitive domains of C4, C5 and C6. Based on Table 7. the results of the assessment of the evaluation question expert stated that the student evaluation instrument was very feasible to use. The average percentage obtained is 96.75% with the criteria "Very Feasible".

Fasibility of STEM-besed learner evaluation tools and Mathematics by expert validators of evaluation questions

Table 7. Percentage of Evaluation Question Expert Validation Result

No	Assessment Aspect Refers to Handayani et al., (2020)	Score	Criteria (%)	Description
	Format			
1	The font type and size are appropriate.	5	100	Very Feasible
2	The numbering system is clear.	5	100	Very Feasible
3	Layout/space arrangement.	5	100	Very Feasible
4	Clarity in the division of material.	5	100	Very Feasible
	Description Test Assessment			
	Material			
1	The material on the questions is in accordance with the indicators.	5	100	Very Feasible
2	Questions match the indicators of the skills being measured.	5	100	Very Feasible
3	The questions are in accordance with the indicators (according to the written test for multiple choice forms).	4	80	Worth
4	The content is appropriate for the grade level.	5	100	Very Feasible
5	The material measured is in accordance with competencies (urgency, relevance, continuity and high daily use).	5	100	Very Feasible
6	Appropriateness of question	4	80	Worth

No	Assessment Aspect Refers to Handayani et al., (2020)	Score	Criteria (%)	Description
	boundaries and answers given.			
7	Coverage of the completeness of the material in the evaluation questions.	4	80	Worth
	Contents			
1	Suitability of questions with content.	5	100	Very Feasible
2	The suitability of the answers to each question developed.	5	100	Very Feasible
3	Suitability of questions with cognitive levels (C4, CS, and C6).	4	80	Worth
4	Questions do not contain SARAPP (Ethnicity, Religion, Race, Intergroup, Pornography and Politics).	5	100	Very Feasible
	Construction			
1	Formulate the question in the form of an interrogative or questioning sentence that demands an elaborated answer.	5	100	Very Feasible
2	There are clear instructions on how to work the instrumen.	4	80	Worth
3	Tables, pictures, graphs, diagrams or the like are meaningful (clear information) or have a relationship with the problem being asked.	5	100	Very Feasible
	Language and Writing			
1	Compliance with correct EYD.	5	100	Very Feasible
2	The formulation of the question sentence is communicative and easy to understand.	5	100	Very Feasible
3	The terms used are easy to understand.	4	80	Worth
4	Question sentences use good and correct language.	5	100	Very Feasible
5	The use of foreign language in the items is correct according to the prevalence used in the field/science of biology.	4	80	Worth
6	Use common language or words (not local language).	5	100	Very Feasible
7	Do not use words/phrases that cause multiple interpretations or misunderstandings.	5	100	Very Feasible
	Average		**96.75**	**Very Feasible**

Feasibility of STEM-Based Student Evaluation Tools by Material Expert Validators

Table 8. Percentage of Material Expert Validation Result

No	Indicator Refers to Aprilianti & Astuti (2020)	Score	Criteria (%)	Category
1	Suitability of questions with KI (Core Competencies).	4	100	Very Good
2	Suitability of questions with KD (Basic Competencies).	4	100	Very Good
3	Clarity of fungi material questions.	4	100	Very Good
4	Suitability of questions with the cognitive domain.	4	100	Very Good
5	Appropriateness of answer key.	4	100	Good
6	There is only one answer key.	4	100	Very Good
7	Suitability of questions with STEM.	3	75	Good
8	Does not use words/phrases that cause double meanings.	3	75	Good
9	Using sentences that are clear and easy to understand.	3	75	Good
10	Question items use standard Indonesian language.	4	100	Very Good
	Average		**92.5**	**Very Good**

Based on Table 8. the results of the material expert validation assessment state that the average percentage obtained is 92.5 %, including the criteria "Very Good" for use.

Feasibility of STEM-Based Student Evaluation Tools by Material Expert Validator

Table 9. Percentage of Media Expert Validation Result

No.	Indicator Refers to Aprilianti & Astuti (2020)	Score	Percentage (%)	Category
1	Appropriateness of media and learning objectives.	4	100	Very Good
2	Suitability of media and learner characteristics	4	100	Very Good
3	Typeface suitability.	4	100	Very Good
4	Appropriateness of font size.	3	75	Good
5	Clarity between text and images.	4	100	Very Good
6	Image clarity.	3	75	Good
7	Appropriate use of color proportion.	3	75%	Good
8	The language used is	3	75%	Good

	communicative.			
9	Has visual appeal which includes color, image, shape and size of letters.	3	75%	Good
	Average		**86.1**	**Very Good**

Based on Table 9. the results of the media expert validation assessment state that the average percentage obtained is 86.1% including the criteria "Very Good" for use.

Table 10. *Cronbach's Alpha* Reliability Analysis Result

Evaluation Criteria		
Reference Value	*Cronbach's Alpha* Value	Conclusion
0,7	0,886	Reliable

Teacher Response

Table 11. Percentage of Teacher Response Result

No	Indicator Refers to Aprilianti & Astuti (2020)	Score	Percentage (%)	Category
1	Suitability of questions with Basic Competencies (KD) and Core Competencies (KI).	5	100	Strongly Agree
2	STEM-based evaluation tools following the material studied.	4	80	Agree
3	STEM-based evaluation tools can improve students' understanding of biological concepts.	4	80	Agree
4	Evaluation tools can hone students' science, technology, engineering and mathematics (STEM) skills.	5	100	Strongly Agree
5	STEM-based evaluation tools make it easier for teachers to evaluate students.	5	100	Agree
6	The material presented in the STEM-based evaluation tool is easy for students to understand.	5	100	Strongly Agree
7	The materials in the STEM-based evaluation instrument are in line with the learning objectives.	5	100	Strongly Agree
8	The material in the STEM-based learning evaluation tool is short, concise and clear.	4	80	Agree
9	Questions are in accordance with the material studied.	4	80	Agree
10	Questions are in line with the learning objectives.	5	100	Strongly Agree
11	STEM-based questions.	5	100	Strongly Agree
12	The language used in the STEM-	5	100	Strongly Agree

No	Indicator Refers to Aprilianti & Astuti (2020)	Score	Percentage (%)	Category
	based evaluation tool is easy to understand and comprehend.			
	Average		**93.3**	**Strongly Agree**

Based on Table 11. the results of the biology subject teacher's response stated that the average percentage obtained was 93.3%, including the criteria "Strongly Agree" to be used.

Student Response

Tabel 12. Percentage of Result Learner Response

No	Indicator Refers to Aprilianti & Astuti (2020)	Score	Percentage (%)	Category
1	The STEM-based evaluation instrument contains clear, concise and complete instructions for using the questions.	4	100	Very Good
2	STEM-based evaluation instruments are in accordance with the material I have learned.	4	100	Very Good
3	STEM-based evaluation instruments can improve my understanding of biology concepts.	4	100	Very Good
4	Evaluation instruments can hone skills in the fields of science, technology, engineering and mathenatics.	4	100	Very Good
5	The material presented in the STEM-based evaluation instrument is easy to understand.	3	75	Good
6	The materials in the STEM-based evaluation instrument are in line with the learning objectives.	4	100	Very Good
7	The material in the STEM-based evaluation instrument is short, concise and clear.	3	75	Good
8	Questions in accordance with the material studied.	3	75	Good
9	Questions in accordance with learning objectives.	3	75	Good
10	The language used in the STEM-based evaluation instrument is easy to understand and comprehend.	3	75	Good
	Average		**87.5**	**Very Good**

Based on Table 12, students' responses' results state that the average percentage obtained is 87.75 %, including the criteria "Very Good" for use. The design stage

begins with determining the form of student evaluation question instruments, followed by the preparation of a lattice of evaluation question instruments. The lattice of student evaluation question instruments is made based on learning objectives using the C4, C5 and C6 cognitive domains. The learner evaluation instrument is in the form of multiple choice, which is objective. If answered correctly, each answer is worth 1 point; otherwise, it is worth zero. Learner evaluation instruments are designed in accordance with the achievement indicators and learning objectives in the 2013 curriculum.

The next stage is the validation of the instrument by 3 experts. The validation stage aims to see the feasibility of the evaluation question instrument product through the content validity test. The instrument is said to be valid if the results of data analysis are in accordance with predetermined criteria. An instrument is said to be valid if the results of the validity of the instrument are in accordance with the criteria, meaning that the test results and predetermined criteria have parallels (Arikunto, 2016). Validation is carried out by validators who are expert lecturers who have the competence to provide an assessment of the product developed to assess the advantages and weaknesses of the product (Sugiyono, 2016).

The feasibility of the product is determined based on the average scores across predefined assessment criteria. The development of this STEM-based student evaluation tool was conducted in several stages, involving validation by evaluation experts, media experts, and feedback from teachers and students regarding the developed product. Based on the evaluation results, the instrument is considered highly feasible for use in the learning process. The average percentage obtained is 96.75 % with the criteria "Very Feasible". The aspects assessed by the evaluation question experts include; the format section, consisting of four aspects, the multiple choice test assessment section, subdivided into the material section, consisting of seven aspects of assessment, the content section consists of four aspects of assessment, the construction section consists of three aspects of assessment, then the language and writing section consists of seven aspects of assessment.

The material expert validation assessment results showed an average score of 92.5 %, which falls within the "Very Good" category. The material experts assessed ten indicators, including the alignment with KI (Core Competencies), the relevance of KD (Basic Competencies), the clarity of the questions, the alignment of the questions with the cognitive domain, the accuracy of the answer key, the presence of a single correct answer, the alignment with STEM components, the use of clear and unambiguous language, the use of sentences that are easy to understand, and the adherence to standard Indonesian language in the questions.

Furthermore, the assessment results from media experts state that the average percentage obtained is 86.1%, including the criteria "Very Good" with several suggestions for improvement. Aspects assessed by media experts on student evaluation tools include the suitability of the media and learning objectives, the suitability of the media and the characteristics of students, the suitability of the font type, the suitability of the font size, the clarity between the text and the image, the clarity of the image, the suitability of the use of color proportions, the language used is communicative, and has visual appeal which includes color, image, shape and font size. In a lesson, media serves as a learning aid that supports the teaching and

learning activities, thereby facilitating the attainment of learning goals (Yani, 2023). Media in learning has an important role, one of which is to overcome the limitations of distance, time and sensory power (Siagian, 2024).

The reliability test is carried out to see the degree of consistency of the instrument and to determine the extent to which the measuring instrument can be trusted and relied upon. The Cronbach's Alpha reliability test result is 0.886, so it is concluded that this instrument is reliable with high reliability criteria. Cronbach's Alpha is a state measure with a value range of 0-1.00 (Tomoliyus & Sunardianta, 2020). The minimum value of the Cronbach's Alpha reliability level is 0.70, According to Fleiss et al., (2004) if the Cronbach's Alpha value is >0.8 then the STEM-based student evaluation tool on fungi material is very reliable.

After validation to experts, it was found that the STEM-based learner evaluation tool product developed had "Very Feasible" criteria from the three experts. In addition to the validation of experts, the responses given by teachers as practical learning are also needed to improve the quality of the developed learner evaluation tools. The teacher's response to the STEM-based learner evaluation tool on fungi material is included in the "Strongly Agree" criteria with the average percentage obtained is 93.3%. After conducting an assessment of biology teachers, researchers conducted a response test to students. Testing was carried out to see the students' response to the evaluation tools developed. The average percentage obtained was 87.5%, including the "Very Good" criteria.

CONCLUSION

Based on the research finding and data analysis, it can be concluded that the STEM-based student evaluation instrumen development is valid, reliable and feasible for use. The assessment by evaluation experts yielded an average score of 96.76%, which falls under the "Very Feasible" category. The evaluation by subject matter experts produced an average percentage of 92.5%, categorized as “Very Good”, while the media experts gave an average score of 86.1%, also falling into the “Very Good” category". Item analysis result showed that 13 questions had very good discriminatory power, 10 were classified as good, 12 as sufficient and 5 were deemed insufficient. Regarding the difficulty level, 27 item were categorized as easy, and 13 as difficult. Teacher response analysis revealed a very positive perception, with an average score of 93.3% across all indicators. Similarly, the analysis of student responses showed an average score of 87.5%. indicating a very positive reception. In conclusion, the STEM-based evaluation instrument is considered suitable for use in the learning process, as it has received positive feedback from both students and teachers, and has met expert criteria for quality and effectiveness.

REFERENCES

Andriani D., Hamdu, G., & Karlimah. (2021). Analysis of Assessment Rubrics Based on Education for Sustainable Development and Systems Thinking in

Primary Schools. *Edukatif: Jurnal Ilmu Pendidikan*, 3(4), 1321-1336 **[*In Indonesian language*]**

Aprilianti, P. P., & Astuti, D. (2020). Development of STEM-based LKPD on Flat Surface Building Material for Grade VIII Junior High School. *Jurnal Pembelajaran Matematika Inovatif*, 3(6), 691-702. **[*In Indonesian language*]**

Arifin, Z. (2016). *Learning Evaluation (Principles, Techniques, and Procedures).* Jakarta: Rosda Karya. **[*In Indonesian language*]**

Arikunto, S. (2012). *Fundamentals of Educational Evaluation.* Jakarta: Rineka Cipta. **[*In Indonesian language*]**

Arikunto, S. (2016). *Research Management*. Jakarta: Rineka Cipta. **[*In Indonesian language*]**

Arikunto, S. (2017). *Development of Research and Program Assessment Instruments*. Yogyakarta: Pustaka Pelajar. **[*In Indonesian language*]**

Febriani, A. Elvia, R. & Handayani, D. (2021). Development of Computer-Based Chemistry Learning Evaluation Tools Using Wondershare Quiz Creator for Buffer Solution Material. *Jurnal Pendidikan dan Ilmu Kimia*, 5(2), 191-197. **[*In Indonesian language*]**

Febriana, R. (2019). *Learning Assessment*. Jakarta Timur: PT Bumi Aksara. **[*In Indonesian language*]**

Fleiss, J. L., Levin, B., & Paik, M. C. (2004). The Measurement of Interrater Agreement. *In Statistical Methods for Rates and Proportions* (pp.598-626). Jhon Wiley & Sons, Inc. https://doi.org/10.1002/0471445428.ch18.

Handayani, S., Minarti, S.U., Rachmawati, D., & Wahyono, H. (2020). STEM-Based Learning Evaluation in Economics Subjects. Malang: PT Literindo Berkah Jaya. **[*In Indonesian language*]**

Herak, R., & Lamanepa, G.H. (2019). Meningkatkan Inovasi Siswa dalam Pembelajaran IPA Melalui STEM. *Jurnal BioEducation,* 4(2), 8-14. **[*In Indonesian language*]**

Idawati, M. L. (2019). Inquiry-Based Authentic Learning in STEM Programs on Students' Scientific Literacy Based on Their Problem-Solving Ability Levels. *Jurnal Pendidikan: Teori, Penelitian, dan Pengembangan*, 4(8), 1024-1029. **[*In Indonesian language*]**

Kristanto, I., Kurniawan E. S., & Maftukhin, A.. (2021). Development of STEM-based Physics Learning Outcome Test Instruments to Measure Students' Scientific Process Skills. *Jurnal Inovasi Pendidikan Sains.* 2(2), 76-81. **[*In Indonesian language*]**

Kurniawan, D. D. (2017). Development of Learning Tools Using Screencast-O-Matic Media for Calculus 2 Course Using the 4-D Thiagarajan Method. *Jurnal Siliwangi*. 3(1), 214-219. **[*In Indonesian language*]**

Majid, A. D. (2015). *Scientific Approach in the Implementation of the 2013 Curriculum.* Bandung: PT. Remaja Rosdakarya. **[*In Indonesian language*]**

Mawardi, K. D. (2021). Development of Research Instruments for Assessing Attitudes Towards Mutual Cooperation in Thematic Learning in Primary Schools. *Jurnal Pendidikan*, 3(3), 640-648. **[*In Indonesian language*]**

Mendiknas. (2003). National Constitution of the Indonesia Republic No. 20 of 2003. *About National Education System.* https://peraturan.bpk.go.id/details/43920/uu-no-20-tahun-2003. Accessed on May 25Th 2025 **[*In Indonesian language*]**

PISA. (2019). *PISA 2018 Results Combined Executive Summaries Volume I, II & III.* https://www.oecd.org/en/publications/pisa-2018-results-volume-_5f07c754-en.html. Accessed on May 25Th 2025 **[*In Indonesian language*]**

PISA. (2023). *PISA 2022 Results (Volume I).* OECD. https://www.oecd.org/en/publications/pisa-2022-results-volume-i_53f23881-en.html. Accessed on May 25Th 2025 **[*In Indonesian language*]**

Purnamasari, A., & Rochmawati. (2015). Development of Information and Communication Technology-Based Learning Evaluation Tools with Wondershare Quiz Creator Inventory Assessment System Material. *Jurnal Pendidikan,* 3(1), 1-9 **[*In Indonesian language*]**

Safitri, R. (2020). *Biology Specialisation in Mathematics and Natural Sciences.* Surakarta: CV Mediatama. **[*In Indonesian language*]**

Siagian, E. M., Listantia, N., Sebayang, Y., Hutagalung, V.k., & Pertiwi, S. (2024). Development of Video Learning Media for Elementary School Studen. *Jurnal Indonesia Berdaya,* 5(2), 621-626.

Simatupang, H., Sianturi, A., & Alwardah, N. (2019). Development of LKPD Based on Science, Technology, Engineering, and Mathematics (STEM) Approach to Foster Students' Critical Thinking Skills. *Jurnal Pelita Pendidikan.* 7(4), 170-177. **[*In Indonesian language*]**

Sudaryono. (2012). *Fundamentals of Learning Evaluation.* Yogyakarta: Graha Ilmu. **[*In Indonesian language*]**

Sudjana, N. (2016). *Assessment of Teaching and Learning Processes.* Bandung; PT Remaja Rosdakarya. **[*In Indonesian language*]**

Sugiyono. (2016). *Statistics for Research*. Bandung: Alfabeta. **[*In Indonesian language*]**

Suparya, I. K. I Wayan Suastra, & Putu Arnyans, I. B. (2022). Low Scientific Literacy: Causes and Alternative Solutions. *Jurnal Ilmiah Pendidikan Citra Bakti*, 9(1), 153-166. **[*In Indonesian language*]**

Thiagarajan, S., Semmel, D.S. & Semmel, M.I.. (1974). *Intructional Development For Training Teachers of Exceptional Children*. Washington, D.C : National Center for Improvement of Educational System.

Tomoliyus, T., & Sunardianta, R. (2020). Validity and Reliability of the Table Tennis Agility Reactive Test Instrument. *Jurnal Keolahragaan,* 8(2), 148-157. **[*In Indonesian language*]**

Yani, F. & Rahmi, L. (2023). Development of Biology Learning Videos on the Main Topic of the Circulatory System for Grade VIII Junior High School/MTs Students for the 2021/2022 Academic Year. *Journal on Education,* 5(4), 12692-12700. **[*In Indonesian language*]**

How To Cite This Article, with *APA style* :

Oetari, N. A., & Idramsa, I. (2025). Development of Science Technology Engineering and Mathematics (STEM) Embedded Evaluation Tool for Micology Learning. *Jurnal Pembelajaran dan Biologi Nukleus*, *11*(2), 769-784. https://doi.org/10.36987/jpbn.v11i2.7235

Conflict of interest : The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author contributions : All authors contributed to the study's conception and design. Material preparation, data collection and analysis were performed by all authors. The first draft of the manuscript was submited by [**Nurul Alfi Oetari**]. All authors contributed on previous version and revisions process of the manuscript. All authors read and approved the final manuscript.