UNIVERSITAS SEBELAS MARET
BIOEDUKASI: JURNAL PENDIDIKAN BIOLOGI
https://jurnal.uns.ac.id/bioedukasi
1693-265X (Print) | 2549-0605 (Online)

The Effect of Heyzine on The Interests and Learning Outcomes of Grade XI Students

Muhammad Iqbal Maulana [a, 1], Muhammad Iqbal Filayani [b, 2, *], Isnawati [a, 3], Desi Kartikasari [b, 4], Nanang Purwanto [b, 5], Aulia Rahma Syita [a, 6], Warda Nadya Paramita [c, 7], Didik Pratikno [c, 8]

[a] Postgraduate Programme in Biology Education, Faculty of Mathematics and Natural Sciences, Surabaya State University, Surabaya, Indonesia
[b] Biology Education, Faculty of Tarbiyah and Teacher Training, Sayyid Ali Rahmatullah State Islamic University, Tulungagung, Indonesia
[c] Biology, MAN 4 Jombang, Indonesia
[1] iqbalmaulanatbio01@gmail.com; [2] muhammadiqbalfilayani16@gmail.com; [3] isnawati@unesa.ac.id;
[4] desikartikasari@uinsatu.ac.id; [5] npurwanto7@gmail.com; [6] 25031465015@mhs.unesa.ac.id;
[7] wardanadyaparamita@gmail.com; [8] azkaavempace@gmail.com
* Corresponding author: muhammadiqbalfilayani16@gmail.com

Submission : 27/12/2025
Revision : 10/08/2026
Accepted : 24/08/2026

ABSTRACT

Teaching at MAN 4 Jombang remains teacher-centered. Adequate facilities have not yet been utilized to their full potential. Survey results indicate that 76.15 percent of students experienced difficulties in learning material on the human musculoskeletal system. This study aimed to determine the effect of a video-based e-magazine on pencak silat on students' interest and learning outcomes in the human musculoskeletal system. This study employed a quantitative method. The research sample comprised Class XII MIPA as a small-scale group (11 students), Class XI MIPA 3 (30 students) as the experimental group, and Class XI MIPA 6 (30 students) as the control group in a large-scale study. Data were collected through questionnaires, observations, tests, and documentation. The results indicated a significant difference in learning outcomes in the experimental class (Wilcoxon sig (2-tailed) = 0.000 < 0.05). The Wilcoxon test for the control class yielded a p-value of 0.325 > 0.05 (no significant difference in learning outcomes). The Mann-Whitney test for learning interest between the experimental and control classes showed a p-value of 0.000 < 0.05 (a significant difference). The average learning interest score for the experimental class was 72.53 percent, compared with 65.1 percent for the control class. This study contributes to innovation in Heyzine-based digital learning rooted in local culture. Further research is directed at measuring students' psychomotor abilities on the human movement system material. The research helps preserve Indonesian silat culture.

Keywords: Heyzine, Interest, Learning Outcomes, Pencak Silat

Introduction

Education is a basic human need. Education enables individuals to understand things they did not know before. Teachers play a vital role in education. As crucial contributors and pioneers, teachers actively help ensure that teaching and learning activities run smoothly. Teachers are responsible for meeting students' needs, so they must have the necessary capabilities and competencies to teach skillfully (Risdiany & Herlambang, 2021). Learning is the ongoing academic communication between teachers and students in the classroom. This means that two human components, namely students and teachers, continuously communicate during the learning process (Suyadi, 2014). Student learning achievement is sometimes related to difficulties in understanding the subject matter during the learning process. These difficulties can stem from ineffective learning characteristics, which can even result in low engagement during class. The negative impact of this is that students fail to learn the material taught by the educator. Teachers find it difficult to meet these students' needs because of these ineffective learning challenges. To motivate students, teachers must create learning innovations that interest them and meet their needs. Students can have an enjoyable learning experience when provided with learning facilities such as media that are relevant to their needs, characteristics, and technological advances.

Using learning media not only helps teachers deliver education to students, but also encourages students to be more actively involved and prompts feedback in the classroom, so teachers and students can receive feedback. Using media during teaching and learning activities also have a very positive impact on learning success (Nurul, 2019). Learning media is an alternative used to reduce misperceptions (Swandi et al., 2015). Anything that can convey a message can attract the attention of the audience, especially students, so that the learning process can run smoothly (Purnawati, 2001). Utami (2012) considers the human musculoskeletal system a difficult biological subject to study. Textbooks sometimes explain the human musculoskeletal system with lengthy explanations, memorization, and boring text and images. Because most of the content is black and white, some animations in human movement material are not very impressive, which can make the content feel boring and lead to misunderstandings. Advances in information technology in education have spurred innovations in learning media that enhance learner engagement and link scientific concepts to real-life phenomena. One such innovation is the use of electronic magazines, which integrate various multimedia elements such as text, images, animations, and videos. Biology teaching at MAN 4 Jombang remains teacher-centered and has not yet made optimal use of educational technology.

A questionnaire distributed to Year 11 students at MAN 4 Jombang showed that 76.15% of students found it difficult to learn about the human musculoskeletal system. A total of 84.55% of students agreed that the media developed should be in the form of videos, and 8.4% agreed that it should be in the form of electronic magazines. Therefore, it is important to create Heyzine-based videos to increase students' interest in and knowledge of human movement material. Magazines published digitally and accessible online are known as electronic magazines. Electronic magazines are a type of educational media designed to stimulate students' interest in reading. Electronic magazines include interesting texts, photos, and videos. Therefore, electronic magazines are designed as e-magazines, which allow students to access them easily at any time and place. Electronic magazines are more efficient than printed magazines, as printed magazines still require high printing costs and cannot include animations, videos, or music. Electronic magazines can save costs, and their file format can be accessed at any time through applications or software on smartphones or personal devices. Electronic magazines are also more environmentally friendly because they are not printed on paper.

In 2014, Jalilah conducted research showing that using electronic magazines combined with flipbooks can improve student learning outcomes. A total of 62.5% of students stated that using electronic magazines combined with flipbooks was very good, and 100% of students passed the minimum competency test with a score of 72, based on the post-test results. Based on the phenomena described above, it is important to research the effect of implementing video-based electronic magazines on pencak silat in human movement system material on the interest and learning outcomes of grade XI students at MAN 4 Jombang. Previous research has not integrated aspects of Indonesian culture into the development of pencak silat-based learning materials. Yet the relationship between pencak silat culture and the subject matter of the human movement system holds educational value that can enrich students' understanding. Integrating pencak silat culture into learning can serve as a strategy for introducing local wisdom whilst enhancing students' ability to link traditional knowledge with scientific concepts.

Methods

This study implements a quantitative approach in the form of pre- and quasi-experiments. Quantitative research focuses on analyzing numerical data, namely figures processed using statistical procedures (Bambang, 2022). A pre-experimental trial design was used for small-scale trials involving only one group and no control class, with a one-shot case study design. A purposive random sampling approach was used. The small-scale trial sample consisted of Class XII MIPA, comprising 11 students. The decision to select 11 students for the small-scale trial (pre-experiment) was based on Achmad (2021), who stated that a small-scale trial consists of 3 to 5 students, whilst a large-scale trial involves 30 or more students. The large-scale trial design used a quasi-experimental approach to investigate the effect of two different treatments on two dependent variables: students' interest and learning outcomes. The sample in the large-scale trial (quasi-experiment) comprised Class XI MIPA 3 (experimental group), consisting of 30 students, and Class XI MIPA 6 (control group), consisting of 30 students. Table 1 shows the pre-experimental research design, and Table 2 shows the quasi-experimental design.

Table 1. Pre-experimental research design in small-scale trials

Class	Number of students	Treatment	*Post-test*	Interest in Learning
XII MIPA	11	A	B	C

Description:
A: Implementation of an electronic magazine (Heyzine) based on pencak silat videos
B: Completion of the post-test for the small-scale trial class
C: Conducting a questionnaire on learning interest in a small-scale trial class

Table 2. Quasi-experimental research design in large-scale trials

Class	Number of students	*Pre-test*	Treatment	*Post-test*	Interest in Learning
XI MIPA 3	30	P1	T1	O1	M1
XI MIPA 6	30	P2	T2	O2	M2

Description:
P1: Completion of the experiment class pretest
P2: Completion of the control class pretest
T1: Implementation of an electronic magazine (Heyzine) based on pencak silat videos

T2: Without the use of Heyzine
O1: Completion of the post-test for the experimental class
O2: Completion of the control class post-test
M1: Filling out the learning interest questionnaire for the experimental class
M2: Filling out the learning interest questionnaire for the control class

Three different types of tests, namely validity, reliability, and prerequisites, are used for data analysis in quantitative research. To assess the quality of the instrument, a validity test is required, with an emphasis on the instrument's ability as a measuring tool. The results of the trial test of respondents who had studied human motion system material, namely 11 students in Grade XII MIPA (N=11), were used for empirical validity testing of the test instrument with a significance level of 5% (0.05), resulting in a table r of 0.602 (Sugiyono, 2013). For the empirical validity test of the learning interest instrument, it was obtained from 5 students (N=5) in the 12th-grade MIPA class at a significance level of 5% (0.05), resulting in a table r of 0.878 (Berliana, 2024).

The next stage is reliability testing, which assesses how accurate or consistent the measurement results are. If the instrument produces accurate data when used by different people at different times, it is considered reliable, accountable, and valid (Himawan & Nurgiyantoro, 2022). Arikunto (2006) states that the reliability index correlation values of the instruments are classified as follows:

Table 3. Interpretation of the reliability correlation index of test questions and learning interest instruments

Correlation Value	Category
0,800 – 1,000	Highly reliable
0,600 – 0,779	Reliable
0,400 – 0,599	Quite reliable
0,200 – 0,399	Somewhat reliable
0,000 – 1,199	Unreliable

After the instruments were tested on a small scale and declared valid and reliable, the next stage was to test them on a large scale. Data collection was carried out by distributing questionnaires, conducting observations and tests, and compiling documentation. Questionnaires were used to collect data on students' interest in learning, tests were used to obtain data on student learning outcomes, and documentation was used as evidence that the research had been carried out. The tests implemented were pre-tests and post-tests. The research was conducted on 13 and 15 April 2025 at MAN 4 Jombang. The researcher used the N-Gain test to identify the diversity of learning outcome data. N-Gain was calculated using SPSS 23. Meltzer's normalized gain test (Latief, 2014) was calculated using the following formula:

$$N\ Gain = \frac{Initial\ test\ score\ - Final\ test\ score}{Ideal\ score - Final\ test\ score}$$

The results of the N-gain calculation are then classified according to the following criteria:

Table 4. Criteria for N Gain levels (Firdaus et al., 2021)

Average	Category
g ≥ 0.7	high
0,7 > g ≥ 0.3	currently
g < 0.3	low

The next step is to test the discriminating power of the questions, using the following formula (Maulana, 2023):

$$KP = \frac{T1}{T2} - \frac{T3}{T4}$$

Description:
KP: Discriminatory capacity
T1: Number of students who answered correctly in the top group
T2: Total number of students who took the test in the upper group
T3: Number of students who answered correctly in the lower group
T4: Total students who took the test in the lower group

The results of the discriminant analysis were then classified according to the following criteria:

Table 5. Discriminant index criteria

No	Discriminant index	Classification
1	0,71 – 1,00	Very good
2	0,41 – 0,70	Good
3	0,21 – 0,40	Bad
4	0,00 – 0,20	Very bad

Source: Arikunto in Maulana, 2023

The next step is to test the data's normality. The purpose of the normality test is to determine whether the sample data originates from a population with a normal distribution. Parametric tests can be used in research if the results indicate that the population is normally distributed. Nonparametric statistical tests can be used if the findings indicate that the population has a non-normal distribution (Tri et al, 2024). Because there are 30 students in each class, this study used the Shapiro-Wilk normality test (Alfiah et al., 2022). To conclude normality testing, it is assumed that the data are normally distributed if the sig value is higher than 0.05 and not normally distributed if the sig value is lower than 0.05 (Kadir, 2017). If the collected data is not normally distributed, then the nonparametric Wilcoxon test is used. The basis for determining whether to accept hypothesis a in the Wilcoxon test is whether the sig value is lower than 0.05. If the Sig value is higher than 0.05, then Ha is rejected (Hakim et al., 2023). The next step is to test data homogeneity. The homogeneity test is conducted using SPSS 23. This test is based on the standard that data are homogeneous if the sig value is higher than 0.05. Data are not homogeneous if the sig level is less than 0.05 (Zahara, 2019). The Mann-Whitney test is a nonparametric statistical test that should be used when the results of the pretest, posttest, N-Gain, and learning interest tests do not appear to be normally distributed and are not homogeneous. The decision regarding the Mann-Whitney U test is based on the assumption that hypothesis a is accepted if the Sig value is lower than 0.05 (Nalakandiarsi, 2019).

Results and Discussion

Before conducting small-scale trials, researchers need to develop and test the feasibility of each research instrument. This helps evaluate the instrument's items. When creating and compiling test questions, researchers must first create a grid. Using a grid can create a test question design that is generally consistent. The first step in compiling a question grid is to identify learning outcomes. Based on the previous material factors, the application of questions to indicators must also be considered (Muntu, 2022). In this study, an expert, Dr Eni Setyowati, S.Pd., MM, a lecturer at UIN Sayyid Ali Rahmatullah Tulungagung, tested the feasibility of the test questions and learning interest questionnaire. Based on the feasibility test results, the test questions received a score of 86%, and the learning interest questionnaire received a score of 90%. It was therefore concluded that the test questions and learning interest questionnaire were very feasible and could be implemented in the research. After education experts tested the test and learning interest instruments for feasibility, students who had studied human movement systems then tested them for validity and reliability in a small-scale trial. The results of the test item validity test are presented in Table 6; the results of the learning interest questionnaire item validity test are presented in Table 7; the results of the test item reliability test are presented in Table 8; and the learning interest questionnaire reliability test is presented in Table 9.

Table 6. Results of test item validity analysis

Description	Question Number	Number
Valid	3, 4, 7 , 15, 18, 24, 28	7 item
Invalid	1, 2, 5, 6, 8, 9, 10, 11, 12, 13, 14, 16, 17, 19, 20, 21, 22, 23, 25, 26, 27, 29, 30	22 item

If a test truly measures what it is supposed to measure, then it is considered valid (Amelia, 2016). A good test question is neither too easy nor too difficult (Amelia, 2016). An instrument item is proven to be valid if the calculated r value is higher than the table r value (Meivira et al., 2022). An instrument item is proven invalid if the measured r value is lower than the table r value. Table 6 shows that seven test items are proven valid, while twenty-two items are proven invalid. The seven valid items will be implemented as pretest and posttest instruments in a large-scale trial, while the twenty-two invalid items will be removed and not used.

Table 7. Validity results of the interest questionnaire trial

Description	Question Number	Number
Valid	1, 3, 4, 5, 9, 11, 12, 18, 19, 20, 21, 23, 24, 25	14 item
Invalid	2, 6, 7, 8, 10, 13, 14, 15, 16, 17, 22	11 item

An instrument item is proven to be valid if the measured r value is higher than the table r value (Meivira et al., 2022). An instrument item is proven invalid if the calculated r value is lower than the table r value. Based on Table 7, 14 learning interest instrument items are valid and 11 are invalid. The researcher used only 12 of the 14 valid learning interest items for the large-scale trial, while the invalid items were eliminated and not implemented.

Table 8. Results of the reliability test of the test instrument

Cronbach's Alpha	N of Items
.803	29

Table 9. Results of the learning interest instrument reliability test

Cronbach's Alpha	N of Items
.975	25

According to Zein 2013, item validity is linked to reliability. The more valid an item is, the higher its reliability. Based on Tables 8 and 9, the test items and learning interest items are highly reliable.

Table 10. Item discrimination index

Number Question	Discriminant Capacity Index	Category
1	0.13	Not good
2	0.6	Good
3	0.33	Not good
4	0	Not good
5	0.87	Very good
6	1	Very good
7	0.067	Not good

The discriminating power of a test is intended to identify differences in student ability. The higher the item index, the better the item, as it can select both gifted and less gifted students. Conversely, the lower the item index, the less successful the item is in distinguishing between capable and less capable students (Amelia, 2016). Based on the results of the identification of discriminating capacity in Table 10, there were two items with very good discriminating capacity, one item with good discriminating capacity, and four items with weak (not good) discriminating capacity.

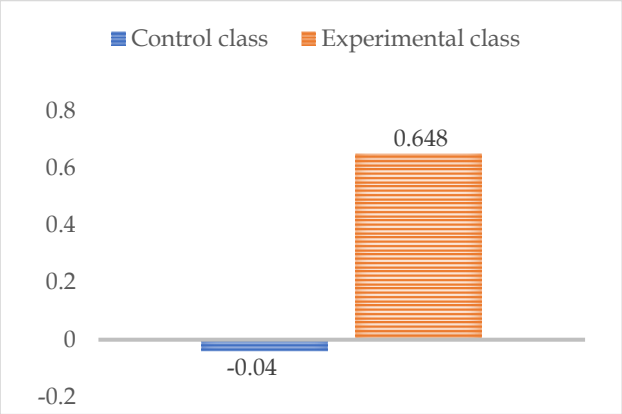

Diagram 1. N Gain learning outcomes

In education, assessment and evaluation are often considered synonymous. Evaluation aims to determine whether learning objectives have been achieved and to improve the learning process. Evaluation is a procedure used to ascertain whether students have achieved learning objectives (Kusumawati & Hadi, 2018). Measurement and assessment are used to achieve learning objectives. Teachers can diagnose students' abilities through measurement and assessment, enabling them to determine the best course of action to improve students' skills (Nurhalimah et al., 2022). Evaluation and assessment are not the same thing. Assessment describes the findings of measurements on a single element, while evaluation determines an element's value and significance through review and decision-making (Arifin, 2016). Based on the assessment and measurement results presented in Diagram 1, the average N Gain value for the control class was -0.04, and the average N Gain value for the experimental class was 0.648.

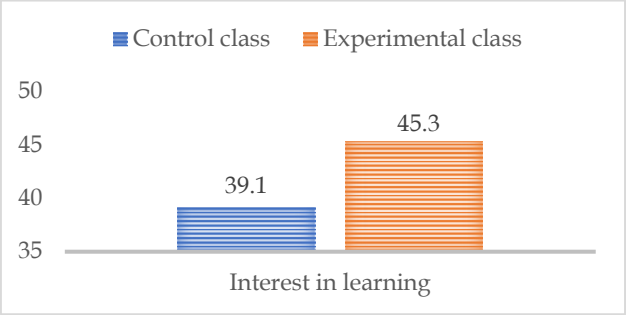

Diagram 2. Learning interest results

Interest is an essential factor in learning activities (Araniri, 2018). Happiness during lessons, sympathy, student participation, and focus are indicators of interest in learning. If students an interested in acquiring knowledge, learning activities will be carried out. Therefore, during the teaching and learning process, teachers must foster students' interest in learning through various methods to produce students who are enthusiastic about learning (Widiati et al., 2022). Based on Diagram 2, the average learning interest score for the control class was 39.1, and for the experimental class it was 45.3.

Table 11. Results of the normality test for N Gain

Class	Shapiro-Wilk		
	Statistic	df	Sig
Control	.697	30	.000
Experiment	.899	30	.008

The normality test in Table 11 shows that the control class has a sig value of 0.000 < 0.05 and the experimental class has a sig value of 0.008 < 0.05, indicating that the N-Gain values of student learning outcomes are not normally distributed. Therefore, a nonparametric test, namely the Wilcoxon test, must be used.

Table 12. Wilcoxon test results

	Posttest control - pretest control	Posttest experiment – pretest experiment
Z	-.984[b]	-4.657 [b]
Asymp. Sig. (2-tailed)	.325	.000

The Wilcoxon test of the control class pretest and posttest data showed a sig value of 0.325 > 0.05, indicating no significant difference in learning outcomes at the beginning and end of the learning process. The Wilcoxon test analysis of the pretest and posttest data for the experimental class showed a significant difference in learning outcomes, as indicated by a sig value of 0.000 < 0.05.

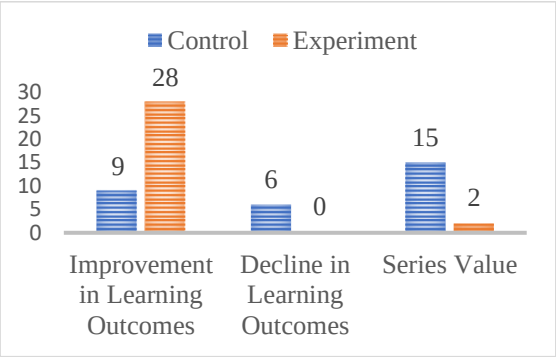

Diagram 3. Number of students experiencing improvement, decline, or stagnation in learning outcomes

Based on Diagram 3, six students experienced a decline in learning outcomes, as indicated by post-test scores lower than their pre-test scores. In the experimental class, no students had post-test scores lower than their pre-test scores. In the experimental class, 28 students had post-test scores higher than their pre-test scores, and in the control class, 9 students had post-test scores higher than their pre-test scores. In the experimental class, 2 students had equal scores, while in the control class, 15 students did. The similarity between the pre-test and post-test scores indicated that the scores were equal.

Table 13. Normality test of learning interest results

Class	*Shapiro-Wilk*		
	Statistic	df	Sig
Control	.968	30	.477
Experiment	.838	30	.000

The Shapiro-Wilk normality test in Table 13 shows that the results of the student learning interest questionnaire in the experimental class have a significance value of 0.000 < 0.05. In contrast, the control class has a sig value of 0.477 > 0.05. Based on these results, the learning interest questionnaire results for the experimental class are not normally distributed, while the results for the control class are normally distributed. Use the Mann-Whitney nonparametric test when the data are not normally distributed.

Table 14. Results of the N Gain homogeneity test

Levene Statistic	df1	df2	Sig.
.712	1	58	.402

Based on Table 14, the N-Gain homogeneity test of the learning outcomes of students in the experimental and control classes produced a Sig value of 0.402. Since the Sig value is greater than 0.05, the N-Gain data of the learning outcomes of students in the experimental and control classes are declared homogeneous, or have identical variability.

Table 15. Homogeneity test results

Levene Statistic	df1	df2	Sig.
1.424	1	58	.238

Based on Table 15, the homogeneity test of the learning interest results of the experimental and control classes produced a Sig value of 0.238. Because the Sig value is greater than 0.05, the learning interest results of the experimental and control classes have similar variability.

Table 16. Mann-Whitney test results for N gain pretest posttest values, learning interest values in control and experimental classes

	Ngain Learning Outcomes	Learning Interest
Mann-Whitney U	.49.000	189.000
Wilcoxon W	514.000	654.000
Z	-6.025	-3.870
Asymp. Sig. (2-tailed)	.000	.000

Since the data in this study were not normally distributed, the hypothesis was tested using the Mann-Whitney test (Nalakandiarsi, 2019). The statistical test obtained a Sig value of 0.000 for N gain in learning outcomes and learning interest, indicating that hypothesis a was accepted and hypothesis 0 was rejected based on the criterion of Sig 0.000 < 0.05. Based on these data, it can be concluded that using a video-based electronic magazine (Heyzine) in

human motion system material has a positive and significant effect on the interest and learning outcomes of Grade XI MIPA MAN 4 Jombang students.

A video-based pencak silat e-magazine (Heyzine) incorporates the concept of cooperative learning, which has the potential to yield several benefits. Firstly, because students can learn and tackle new challenges both individually and in groups, this activity enables them to process the material more thoroughly, leading to better results as they enjoy exchanging ideas with one another while working in groups. Learning activities in the control class, which used a conventional approach, showed that students were dozing off and not paying attention to the teacher whilst they were explaining. Students in the control class were only active in asking questions during the question-and-answer session. The lack of varied learning media caused students to become bored and uninterested in the material taught by the teacher. Monotonous teaching can hinder students' intrinsic motivation, as it typically involves one-way communication between the teacher and students without opportunities for interaction or student engagement. Students may become less motivated to learn and pursue a deeper understanding of the subject matter if they do not feel actively involved in the learning process. Students in the control class (XI MIPA 6) became less motivated to learn, which may have an impact on their academic performance and cognitive development. Here is the link to the video-based pencak silat e-magazine (Heyzine):
https://canva.link/vul0j0sf40vlgbw
https://youtu.be/mSSjNMkHf5g?si=ItQsDkqEfdEuNm0w

Contributions of the research and future research directions

This research contributes to introducing and instilling a sense of pride in pencak silat as part of Indonesia's cultural heritage amongst the younger generation through the study of biology. Pencak silat is not merely a sporting activity or a form of self-defense, but also part of the nation's identity and culture that can be studied from a scientific perspective. The use of pencak silat in teaching materials is expected to foster students' awareness that Indonesian culture holds value that can be studied across various fields of science. This can encourage a sense of belonging and students' concern for the sustainability of the nation's culture, ensuring that learning focuses not only on academic achievement but also on the development of national character. Future research should implement hands-on practice of pencak silat movements by combining biology lessons with outdoor physical education outside the classroom, so pupils do not rely solely on audiovisual materials to understand the subject matter but can also gain a kinesthetic understanding. Further research will measure students' psychomotor abilities in the human movement system material.

Conclusion

The results of the study indicate that the scientific explanation skills of students at a high school in Tasikmalaya remain in the low category across all components of claim, evidence, reasoning, and rebuttal as reflected in their limited ability to link empirical evidence with biological concepts to construct logical and structured scientific arguments. This suggests that biology instruction is still dominated by rote memorization rather than conceptual understanding, failing to provide learning experiences that foster scientific thinking skills, particularly in topics like the circulatory system, which require deeper analysis of biological phenomena. Based on the findings of this study, it is recommended that future research develop and test scientific activity-based learning approaches, such as inquiry-based and argument-driven inquiry, that provide students with broader opportunities to explore, observe, interpret data, and construct evidence-based arguments. This study has limitations regarding the sample size and scope; therefore, further research with a larger, more representative sample is needed. Further research should also consider integrating technology into biology education, such as computational thinking and virtual laboratories, as innovative

approaches that can facilitate simulations of biological processes and support independent, data-driven problem-solving. Developing and testing these learning strategies is expected to further inform the design of interactive, relevant biology instruction, thereby more effectively enhancing students' scientific explanation skills and science literacy.

Acknowledgement

We would like to express our gratitude to MAN 4 Jombang for allowing the research to run smoothly and to all parties who supported the writing of this article and helped complete it successfully.

Generative AI Disclosure

The use of artificial intelligence in this manuscript is limited to language, translation, and editing to improve readability, without affecting the scientific substance, data analysis, or research interpretation.

References

Achmad Noor Fatirul, D. A. W. (2021). *Metode Penelitian Pengembangan Bidang Pembelajaran (Edisi Khusus Mahasiswa Pendidikan dan Pendidik)*. Pascal Books.

Alfiah, S., & Dwikoranto, D. (2022). Penerapan model problem-based learning berbantuan laboratorium virtual phet untuk meningkatkan HOTs siswa SMA. *Jurnal Penelitian Pembelajaran Fisika, 13*(1), 9–18. https://doi.org/10.26877/jp2f.v13i1.11494

Amelia, M. A. (2016). Analisis soal tes hasil belajar high order thinking skills (HOTS) matematika materi pecahan untuk kelas 5 Sekolah Dasar. *Jurnal Penelitian, 20*, 123–131.

Araniri, N. (2018). Kompetensi profesional guru agama dalam menumbuhkan minat belajar siswa. Risalah, *Jurnal Pendidikan Dan Studi Islam, 4*(1), 80. https://doi.org/10.5281/zenodo.3552011

Arifin, Z. (2016). *Evaluasi Pembelajaran*. Rodaskarya.

Arikunto, S. (2006). *Prosedur Penelitian Suatu Pendekatan Praktek (Edisi Revisi VI)*. In PT Rineka Cipta.

Bambang Sudaryana, H. R. R. A. (2022). *Metodologi Penelitian Kuantitatif*. Deepublish.

Berliana, A. O. M. (2024). Analisis uji coba validasi dan reliabilitas instrumen validasi pengembang bahan ajar flipbook dalam pembelajaran berdiferensiasi. *EL-Muhbib Jurnal Pemikiran dan Penelitian Pendidikan Dasar, 8*, 1–12. https://ejournal.iaimbima.ac.id/index.php/eL-Muhbib/article/view/2266%0Ahttps://ejournal.iaimbima.ac.id/index.php/eL-Muhbib/article/download/2266/1087

Firdaus, R. J., Wahyuni, S., & Utomo, A. P. (2021). Analisis penggunaan video pembelajaran IPA kontekstual terhadap kemampuan berpikir kritis siswa SMP. *Physics and Science Education Journal (PSEJ), 1*, 50–56. https://doi.org/10.30631/psej.v1i2.702

Hakim, A. R. Al, Imron, A., Suprijono, A., & Marzuqi, M. I. (2023). Pengaruh penggunaan media pembelajaran tumpeng nasional terhadap hasil belajar IPS siswa kelas VIII. *Ejournal.Unesa.Ac.Id, 3*(1), 25–35. https://ejournal.unesa.ac.id/index.php/PENIPS/article/view/51059%0Ahttps://ejournal.unesa.ac.id/index.php/PENIPS/article/view/51059/42255

Himawan, R., & Nurgiyantoro, B. (2022). Analisis butir soal latihan penilaian akhir semester ganjil mata pelajaran bahasa Indonesia kelas VIII SMPN 1 Bambanglipuro Bantul

menggunakan program ITEMAN. *KEMBARA Journal of Scientific Language Literature and Teaching*, *8*(1), 160–180. https://doi.org/10.22219/kembara.v8i1.20530

Jalilah, R. N., & Sukarmin, Teguh Rahardjo, D. (2014). Pengembangan media pembelajaran interaktif e-magazine pada materi pokok dinamika rotasi untuk SMA kelas XI. *Jurnal Materi Dan Pembelajaran Fisika (JMPF)*, *4*(1), 18–25. https://doi.org/https://studylibid.com/doc/957140/jurnal-materi-dan-pembelajaran-fisika--jmpf-

Kadir. (2017). Statistika terapan: konsep, contoh dan analisis data dengan program SPSS/Lisrel dalam penelitian. 147–148. https://api.semanticscholar.org/CorpusID:64505726

Kusumawati, M., & Hadi, S. (2018). An analysis of multiple-choice questions (MCQs): Item and test statistics from mathematics assessments in senior high school. *REID (Research and Evaluation in Education)*, *4*(1), 70–78. https://doi.org/10.21831/reid.v4i1.20202

Latief, H. (2014). Pengaruh pembelajaran kontekstual terhadap hasil belajar. *Jurnal Geografi Gea*, *14*(1), 19. https://doi.org/10.17509/gea.v14i1.3358

Maulana, H. (2023). Analisis kualitas instrumen evaluasi pembelajaran menggunakan media digitalisasi untuk memotivasi hasil belajar peserta didik. *Jurnal Pendidikan Bhinneka Tunggal Ika*, *1*(4), 1–12. https://doi.org/10.51903/bersatu.v1i4.255

Meivira, A., Amelia, N. M., & Puspitasari, C. (2022). Uji validitas dan reliabilitas kuesioner penggunaan dan penyimpanan antibiotika di Kecamatan Ampenan. Archives Pharmacia, 4(January), 10–18. https://unram.sgp1.digitaloceanspaces.com/simlitabmas/kinerja/penelitian/jurnal/199003202018032001-1657011105-uji_validitas_dan_realibilitas_audra.pdf

Muntu, N. (2022). Evaluasi pembelajaran PAI dan budi pekerti pada MTs dan SMP di Kota Bitung. *Journal of Islamic Education Policy*, *7*(2), 135. https://doi.org/http://dx.doi.org/10.30984/jiep.v7i2.2220

Nalakandiarsi, C. (2019). Pengaruh 3 teknik finger painting terhadap kemampuan fisik motorik halus pada anak kelompok B di TK Negeri Pembina Surabaya. *Jurnal PAUD Teratai*, *8*((1)), 1–5. https://doi.org/https://ejournal.unesa.ac.id/index.php/paud-teratai/article/view/28498

Nurhalimah, S., Hidayati, Y., Rosidi, I., & Hadi, W. P. (2022). Hubungan antara validitas item dengan daya pembeda dan tingkat kesukaran soal pilihan ganda pas. *Natural Science Education Research*, *4*(3), 249–257. https://doi.org/10.21107/nser.v4i3.8682

Nurul, A. (2019). Peran media pembelajaran meningkatkan hasil belajar. *Posiding Seminar Nasional Pendidikan FKIP*, *2*(1), 586–595. https://doi.org/https://jurnal.untirta.ac.id/index.php/psnp/article/view/5665

Purnawati, E. (2001). *Media pembelajaran*. CV. Rajawali.

Risdiany, H., & Herlambang, Y. T. (2021). Pengembangan profesionalisme guru dalam mewujudkan kualitas pendidikan di Indonesia. Edukatif: *Jurnal Ilmu Pendidikan*, *3*(3), 817–822. https://doi.org/https://doi.org/10.31004/edukatif.v3i3.434

Sugiyono. (2013). *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. In Alvabeta. CV. https://www.academia.edu/118903676/Metode_Penelitian_Kuantitatif_Kualitatif_dan_R_and_D_Prof_Sugiono

Suyadi. (2014). Pelaksanaan pembelajaran Pendidikan Agama Islam (PAI) di SMK Negeri 1 Lais Kecamatan Lais Kabupaten Musi Banyuasin. *Conciencia*, *14*(1), 46. http://jurnal.radenfatah.ac.id/index.php/conciencia/article/view/87

Swandi, A., Nurul Hidayah, S., & Irsan, L. J. (2015). Pengembangan media pembelajaran laboratorium virtual untuk mengatasi miskonsepsi pada materi Fisika Inti di SMAN 1 Binamu, Jeneponto (Halaman 20 s.d. 24). *Jurnal Fisika Indonesia, 18*(52), 20–24. https://doi.org/10.22146/jfi.24399

Tri Wulansari & Muhammad Iqbal Filayani. (2024). The influence of PBL-based worksheets on interest and learning outcomes. *Bioeduca: Journal of Biology Education, 6*(2), 109–124. https://doi.org/10.21580/bioeduca.v6i2.20775

Utami, N. R., Ika Mustikasari, S. (2012). Efektivitas pemanfaatan macromedia flash dengan pendekatan SAVI materi sistem gerak di SMAN 1 Kajen. *Unnes Journal of Biology Education, 1*(2), 102–108. https://www.semanticscholar.org/paper/EFEKTIVITAS-PEMANFAATAN-MACROMEDIA-FLASH-DENGAN-DI-Mustikasari-Rahayu/cb443748a79b563e3bbb43fc0ccc2b7b0f52aad9

Widiati, Sridana, N., Kurniati, N., & Amrullah, A. (2022). Pengaruh minat belajar dan kebiasaan belajar terhadap prestasi belajar Matematika. *Griya Journal of Mathematics Education and Application, 2*(4), 890. https://doi.org/10.29303/griya.v2i4.240

Zahara, H. M. (2019). Pengaruh media papan penjumlahan terhadap hasil belajar matematika materi penjumlahan siswa kelas I Sekolah Dasar. *Jurnal Penelitian Pendidikan Guru Sekolah Dasar*, 2941–2950. https://doi.org/https://ejournal.unesa.ac.id/index.php/jurnal-penelitian-pgsd/article/view/28051

Zein, A., Fadillah, M., & Novianti, R. (2013). Hubungan antara validitas butir, reliabilitas, tingkat kesukaran dan daya pembeda soal ujian semester genap bidang studi Biologi kelas XI SMA / MA Negeri di Kota Padang tahun pelajaran 2010/2011. *Semirata 2013 FMIPA Unila*, 2009, 46. https://jurnal.fmipa.unila.ac.id/semirata/article/download/572/392