

Beyond Surface Learning: Deep Learning-Based Card Sort Media for Enhancing Higher-Order Thinking Skills in Biology Education

Asdar Efendi

Master's Program in Biology Education, Postgraduate Faculty, Universitas Patompo, Makassar, Indonesia

Muhammad Sri Yusal

Master's Program in Biology Education, Postgraduate Faculty, Universitas Patompo, Makassar, Indonesia

Herlina

Master's Program in Biology Education, Postgraduate Faculty, Universitas Patompo, Makassar, Indonesia

Corresponding Author: herlinamuh.tahir@unpatompo.ac.id

Abstract

Twenty-first-century education calls for a shift from surface, rote memorisation toward deep, higher-order thinking, yet many active-learning media, including the popular Card Sort technique, still risk facilitating only surface engagement. This study examined a Deep Learning-based Card Sort medium designed to trigger cause-effect reasoning rather than visual matching, testing its effect on learning activity and learning outcomes using a quasi-experimental, Nonequivalent Control Group design at a senior secondary school in Indonesia, on the human movement system topic. The sample comprised an experimental class of 22 students and a control class of 25 students. Data were collected through a learning-activity observation sheet and pretest-posttest achievement tests, and analysed with descriptive statistics, N-Gain, normality and homogeneity tests, and an independent samples t-test. The experimental class recorded a mean learning-activity score of 86.73% (very active) against 75.56% (active) in the control class. Its posttest mean reached 90.55 with 100% mastery learning, compared with 82.08 and 84% mastery in the control class, and its mean N-Gain was 0.7396 (high category) against 0.5533 (moderate) in the control class. The independent samples t-test confirmed a significant difference between groups ($t = 3.9221$, $p = 0.0003$). These findings indicate that embedding deep-learning

1186



principles into Card Sort media meaningfully strengthens both engagement and higher-order achievement, offering a low-cost, globally transferable model for active learning reform in science classrooms.

Keywords: Card Sort; Deep Learning; Higher-Order Thinking Skills; Learning Activity; Learning Outcomes.

INTRODUCTION

A central challenge facing science education worldwide is the persistent gap between curricular ideals and classroom practice: policy increasingly calls for deep, transferable understanding and higher-order thinking skills, yet instruction in many classrooms remains dominated by lecturing and rote memorisation, leaving learners able to recall facts without being able to reason about them. Indonesia's Merdeka Curriculum reflects this global shift, reorienting learning away from content coverage toward problem-solving and deep conceptual understanding, but the gap between this ideal and everyday practice remains wide, particularly in Biology, where topics such as the human movement system appear deceptively simple yet demand genuine conceptual, cause-effect reasoning rather than memorisation.

Active-learning strategies such as Card Sort, in which students physically sort and discuss cards to build categories, are widely used to counter this passivity (Silberman, 1996). However, comparative evidence shows the strategy's limits: Situmorang (2020) found card-sort instruction significantly outperformed conventional teaching on an ecosystem topic, while a broader review by Lubis et al. (2024) confirmed consistent gains in activity and outcomes across studies, yet Pratama and Jannah (2023) and Rahayu et al. (2025) both caution that, without redesign, students often merely match cards by surface visual similarity rather than genuinely reasoning about the concepts involved – a risk long identified in Marton and Säljö's (1976) classic distinction between Surface Learning, driven by intent to complete a task, and Deep Learning, driven by intent to understand.

This study addresses that gap by developing and testing a Deep Learning-based Card Sort medium, in which cards function as problem triggers built around cause-effect relationships rather than rote matching, drawing on evidence that well-designed problem-based card mechanics can trigger genuine insight (Douchy, 2023) and that Deep Learning is best understood as active, meaningful, transferable engagement with content rather than a synonym for machine



learning (Haq & Prasetyo, 2025). The medium was trialled at a senior secondary school in Indonesia where preliminary observation indicated low activity and below-target outcomes on the movement-system topic. Specifically, this study examined the effect of Deep Learning-based Card Sort media on (1) students' learning activity, (2) their learning outcomes, and (3) the difference in both measures relative to conventional instruction – questions whose answers speak to a challenge common to science classrooms far beyond this single school.

LITERATURE REVIEW

Deep Learning and Higher-Order Thinking

Instruction is the interaction between educator, learner, and learning resources that helps a learner move from not knowing to knowing (Azani & Sarmila, 2024). Overcoming the complexity of topics such as the movement system requires a shift from Surface to Deep Learning: Surface Learning is characterised by an intent merely to complete a task, manifested as memorising isolated facts, whereas Deep Learning is driven by an intrinsic intent to understand, actively searching for patterns and logical relationships among concepts (Marton & Säljö, 1976). Without explicit instructional support toward Deep Learning, students often fail to develop an expert-like cognitive framework and remain trapped in shallow understanding (Smith, 2017); learning outcomes worth measuring must therefore go beyond remembering (C1) and understanding (C2) toward the higher-order cognitive domain of analysing (C4), evaluating (C5), and creating (C6) as defined in the revised Bloom's Taxonomy (Anderson & Krathwohl, 2001).

Card Sort as an Active Learning Medium

Card Sort is a collaborative active-learning strategy in which students move around the classroom to find cards matching a given category and then justify their grouping (Silberman, 1996). Its strength lies in kinaesthetic and social engagement, but recent literature cautions that, in its traditional form, students often match cards by surface visual similarity without truly discussing the underlying concept (Pratama & Jannah, 2023; Rahayu et al., 2025). This study addresses that limitation by turning each card into a problem trigger built around a cause-effect scenario rather than a simple matching cue, consistent with evidence that problem-embedded card mechanics can generate genuine insight (Douchy, 2023) and with Fullan, Quinn, and McEachen's (2018) view of Deep Learning as the cultivation of global competencies – character,



citizenship, collaboration, communication, creativity, and critical thinking – rather than passive content transfer.

Learning Activity, Learning Outcomes, and Hypotheses

Learning activity should not be equated with physical movement alone: following Diedrich's classification (in Sardiman, 2011), it spans visual, oral, motor, and – most importantly – mental activity, the last reflecting genuine analytical engagement rather than the speed of task completion. Learning outcomes, in turn, are durable changes in cognitive ability; when the intervention is deep-learning oriented, outcomes should be assessed at the higher-order levels (C4–C6) so that score gains reflect genuine conceptual understanding rather than short-term retention. Meaningful learning theory (Ausubel, in Novak, 2010) further predicts that connecting new information to existing knowledge, as a cause-effect card task requires, produces more durable understanding than rote association, while a large-scale meta-analysis across science, engineering, and mathematics found active-learning approaches to significantly outperform conventional lecturing on measured outcomes (Freeman et al., 2014).

Based on this framework, three hypotheses were formulated. First, regarding learning activity: H0 – there is no effect of Deep Learning-based Card Sort media on students' learning activity; H1 – there is such an effect. Second, regarding learning outcomes: H0 – there is no effect of the medium on students' learning outcomes; H2 – there is such an effect. Third, regarding the difference between groups: H0 – there is no significant difference in learning activity and learning outcomes between students taught with the medium and those taught conventionally; H3 – there is a significant difference.

METHOD

Research Design

This study used a quantitative, Quasi-Experimental method with a Nonequivalent Control Group design. Quasi-experimentation was chosen because true random assignment of individual students is rarely feasible in a formal school setting without disrupting existing class structures (Campbell & Stanley, 1963); both groups therefore retained their intact classes but received a pretest to establish initial ability before treatment, which is essential for internal validity (Sugiyono, 2019). The design is shown in Table 1.



Table 1. Nonequivalent Control Group Research Design

Group	Pretest	Treatment	Posttest
Experimental (E)	O1	X	O2
Control (K)	O3	–	O4

E is the group taught with Deep Learning-based Card Sort media and K the group taught conventionally; O1/O3 are baseline pretest scores, X is the Card Sort intervention, and O2/O4 are posttest scores used to assess its impact on higher-order learning outcomes (C4–C6).

Population, Sample, and Instruments

The population comprised all grade XI students at the research site (two intact classes) in the 2025/2026 academic year, all following the same curriculum and material. Purposive sampling yielded an experimental class of 22 students and a control class of 25 students. Two instruments were used: a non-test observation sheet recording learning activity on a four-point Likert scale across visual, oral, motor, and mental indicators, completed by two observers for inter-rater objectivity; and a structured essay achievement test targeting the C4–C6 cognitive levels, validated for content by subject-matter experts and for empirical validity via Pearson product-moment correlation on a non-sample try-out, with reliability assessed using Cronbach's Alpha ($r_{11} > 0.60$ considered acceptable; Azwar, 2015).

Procedure and Data Analysis

Data collection proceeded in three stages: preparation (permits, preliminary observation, lesson-module and instrument validation, sample selection), implementation (simultaneous pretest, treatment – Deep Learning-based Card Sort in the experimental class and conventional instruction in the control class – across at least three meetings with continuous activity observation), and completion (posttest and documentation). All data were analysed in IBM SPSS Statistics 27.0 using descriptive statistics and the Normalized Gain (N-Gain; Hake, 1999), categorised as high ($g \geq 0.7$), moderate ($0.3 \leq g < 0.7$), or low ($g < 0.3$). Prior to hypothesis testing, normality was assessed with the Shapiro-Wilk test, chosen for its greater sensitivity with small samples ($N < 50$; Razali & Wah, 2011), and homogeneity with Levene's Test, both at $\alpha = 0.05$. Where these assumptions were met, hypotheses were tested with an independent samples t-



test ($\alpha = 0.05$); the non-parametric Mann-Whitney U test served as an alternative had normality not been satisfied.

RESULTS AND DISCUSSION

Learning Activity

Learning activity data, recorded by peer-teacher observers throughout instruction, are summarised in Table 2. The experimental class achieved a mean activity score of 86.73% (very active category), compared with 75.56% (active) in the control class, a gap illustrated in Figure 1.

Table 2. Recapitulation of Students' Learning Activity

Activity Category	Range (%)	Experimental Class	Control Class
Very Inactive	0% – 20%	0 (0%)	0 (0%)
Inactive	21% – 40%	0 (0%)	1 (4%)
Fairly Active	41% – 60%	0 (0%)	0 (0%)
Active	61% – 80%	6 (27.27%)	17 (68%)
Very Active	81% – 100%	16 (72.73%)	7 (28%)
Mean Percentage	—	86.73%	75.56%

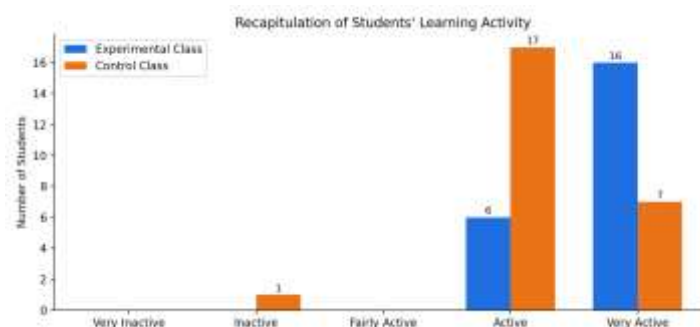


Figure 1. Recapitulation of Students' Learning Activity

Learning Outcomes

Table 3 summarises pretest and posttest scores against the school's minimum mastery criterion of 75. Pretest means were comparable (63.91 vs 58.60), but after treatment the experimental class posttest mean reached 90.55 with 100% mastery learning, against 82.08 and 84% mastery in the control class (Figure 2) – a mean gain of 26.64 points versus 23.48 points.

1191

Table 3. Descriptive Statistics of Learning Outcomes

Statistic	Exp. Pretest	Exp. Posttest	Ctrl. Pretest	Ctrl. Posttest
N	22	22	25	25
Mean	63.91	90.55	58.60	82.08
Std. Deviation	12.84	7.04	9.30	7.68
Minimum	35	75	40	60
Maximum	80	100	75	95

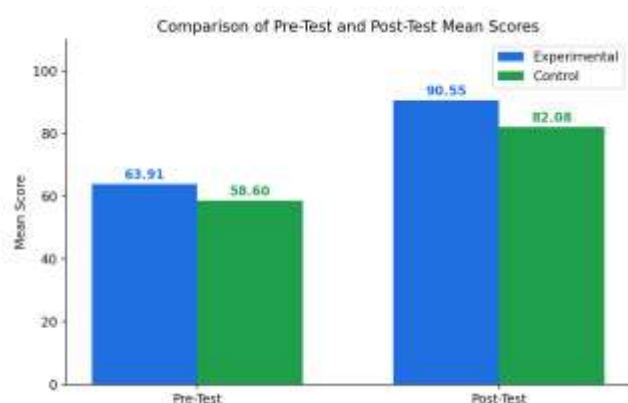


Figure 2. Comparison of Pre-Test and Post-Test Mean Scores

The N-Gain analysis (Table 4) confirms this pattern: the experimental class averaged 0.7396 (high category; 16 of 22 students high, 6 moderate), against 0.5533 (moderate category) in the control class (3 high, 20 moderate, 2 low), indicating that the medium benefited students across ability levels, not only high achievers.

Table 4. Comparison of N-Gain, Experimental and Control Classes

Class	Mean N-Gain	Category	High	Moderate	Low
Experimental	0.7396	High	16	6	0
Control	0.5533	Moderate	3	20	2

Prerequisite and Hypothesis Testing

Both the Shapiro-Wilk normality test (all p-values between 0.065 and 0.193, all > 0.05) and Levene's homogeneity test (pretest $F = 2.89$, $p = 0.096$; posttest $F = 0.03$, $p = 0.873$, both > 0.05) confirmed that the data were normally distributed and the group variances homogeneous, justifying the use of an independent samples t-test for hypothesis testing, the results of which are presented in Table 5.

Table 5. Independent Samples t-Test Results

Comparison	Exp. Mean	Ctrl. Mean	t statistic	p-value	Conclusion
Posttest	90.55	82.08	3.9221	0.0003	Significant

The t-test produced $t = 3.9221$ with $p = 0.0003$ ($p < 0.05$), so H_0 is rejected and H_1 accepted: there is a significant difference in learning outcomes between the experimental and control classes, confirming that Deep Learning-based Card Sort media has a significant effect on both learning activity and learning outcomes.

DISCUSSION

The higher activity recorded in the experimental class is consistent with the medium's integration of physical, mental, and social engagement: as students read, discussed, classified, and presented their cards, they moved through visual, oral, motor, and mental activity in one cohesive sequence, echoing Djamarah and Zain's (2020) account of learning activity as a multidimensional construct and Lubis et al.'s (2024) finding that Card Sort consistently strengthens activity and communicative interaction relative to passive instruction. Framing the cards as cause-effect problems rather than matching cues appears central to this gain: it required students to move beyond visual pattern-matching toward the kind of intent-driven meaning-making that defines Deep Learning (Marton & Säljö, 1976), consistent with Douchy's (2023) argument that problem-embedded card mechanics can trigger genuine insight.

The corresponding gain in learning outcomes – a posttest advantage of 8.5 points and an N-Gain nearly 0.19 higher than the control class – is consistent with meaningful-learning theory (Ausubel, in Novak, 2010): connecting cards through conceptual logic rather than rote association likely helped students integrate new information into existing knowledge, producing more durable understanding, consistent with Fullan, Quinn, and McEachen's (2018) view of Deep



Learning as the cultivation of critical thinking rather than passive content transfer, and with broader evidence that active learning outperforms lecturing across science disciplines (Freeman et al., 2014). That gains were concentrated among initially lower- and moderate-scoring students suggests the medium's problem-trigger design is genuinely instructive rather than merely engaging – a pattern with plausible relevance to any classroom, in any national curriculum, seeking to convert active-learning enthusiasm into measurable higher-order gains.

CONCLUSION

Deep Learning-based Card Sort media had a significant effect on both students' learning activity (86.73% very active versus 75.56% active) and learning outcomes (posttest mean 90.55 with 100% mastery versus 82.08 and 84% mastery, mean N-Gain 0.7396 high versus 0.5533 moderate), with an independent samples t-test confirming a significant difference between the experimental and control classes ($t = 3.9221$, $p = 0.0003$). By redesigning cards as cause-effect problem triggers rather than simple matching cues, the medium engaged students visually, orally, motorically, and mentally more thoroughly than conventional instruction, and translated that engagement into higher-order gains (C4–C6) across ability levels rather than only among high achievers. These findings support Deep Learning-based Card Sort as an effective, low-cost, and readily transferable strategy for advancing higher-order thinking in science classrooms confronting the same surface-to-deep learning gap documented internationally.

RECOMMENDATIONS

Teachers are encouraged to adopt Deep Learning-based Card Sort for conceptual, cause-effect material, ensuring cards embed genuine problem triggers rather than visual matching and pairing them with adequate scaffolding for initially passive students. Schools are encouraged to support wider adoption across science subjects through targeted teacher training. Future research should test the medium on other topics and curricula, employ a larger, fully randomised sample, and explore its integration with digital or gamified tools.

REFERENCES

Anderson, L. W., & Krathwohl, D. R. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman.



- Ausubel, D. P. (in Novak, J. D., 2010). Learning, creating, and using knowledge: Concept maps as facilitative tools in schools and corporations. Routledge.
- Azani, A., & Sarmila, G. (2024). Hakikat belajar dan pembelajaran. Mutiara: Jurnal Penelitian dan Karya Ilmiah, 2(3), 17–37.
- Azwar, S. (2015). Reliabilitas dan validitas. Pustaka Pelajar.
- Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. Houghton Mifflin Company.
- Douchy, T. (2023). Gamification and insight: Problem solving mechanics in classroom. Journal of Educational Psychology.
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. Proceedings of the National Academy of Sciences, 111(23), 8410–8415. <https://doi.org/10.1073/pnas.1319030111>
- Fullan, M., Quinn, J., & McEachen, J. (2018). Deep learning: Engage the world change the world. Corwin Press.
- Hake, R. R. (1999). Analyzing change/gain scores. American Educational Research Association's Division D Measurement and Research Methodology.
- Haq, M. D., & Prasetyo, N. T. (2025). Deep learning sebagai pendekatan transformasional dalam pendidikan: Sebuah tinjauan literatur. Jurnal Studi Guru dan Pembelajaran, 8(3), 1826–1842.
- Lubis, D. C., Ritonga, A. A., Febriani, A., Jannah, M., Syahfitri, N., & Yusnaldi, E. (2024). Studi literatur review: Pengaruh penggunaan media card sort terhadap hasil belajar siswa pada pembelajaran IPS di SD/MI. Jurnal Pendidikan Tambusai, 8(1), 1434–1445. <https://doi.org/10.31004/jptam.v8i1.12592>
- Marton, F., & Säljö, R. (1976). On qualitative differences in learning: I—Outcome and process. British Journal of Educational Psychology, 46(1), 4–11.
- Pratama, Y. A., & Jannah, M. (2023). Pengaruh model active learning tipe card sort terhadap hasil belajar IPA. Jurnal Pendidikan Biologi dan Sains.
- Rahayu, R., Sukardi, S., & Sabrina, R. N. (2025). Analisis efektivitas card sort konvensional pada pembelajaran biologi SMA. Jurnal Pendidikan Biologi.



- Razali, N. M., & Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, 2(1), 21–33.
- Sardiman, A. M. (2011). *Interaksi dan motivasi belajar mengajar*. Rajawali Pers.
- Silberman, M. L. (1996). *Active learning: 101 strategies to teach any subject*. Allyn and Bacon.
- Situmorang, M. V. (2020). Pengaruh media pembelajaran “Card Sort” terhadap hasil belajar Biologi. *Prosiding Seminar Nasional Sains (SINASIS)*, 1(1), 659–667. <https://proceeding.unindra.ac.id/index.php/sinasis/article/view/4044>
- Smith. (2017). From surface to deep: Biology student's cognitive framework. *International Journal of Science Education*.
- Sugiyono. (2019). *Metode penelitian kuantitatif, kualitatif, dan R&D*. Alfabeta.

