

Effectiveness of a teachable machine-based instructional program for developing electronic concepts among eighth-grade students

Mahmoud M. F. Barghot ¹ * , Ahmed Mohamed Abu Elba ² 

¹ University College of Science and Technology, Palestine.

² Ministry of Education and Higher Education, Palestine.

Suggested citation: Barghot, M. M. F. & Abu Elba, A. M. (2026). Effectiveness of a teachable machine-based instructional program for developing electronic concepts among eighth-grade students. *Journal of Educational Technology and Online Learning*, 9(3), 365-373.

Highlights	Abstract
• A Teachable Machine and PictoBlox instructional sequence was designed to develop eighth-grade students' electronic concepts. • A quasi-experimental, nonequivalent pretest-posttest control-group design compared two intact classes (N = 80). • The experimental group substantially outperformed the control group on the post-test, $t(78) = 38.44$, $p < .001$, Cohen's $d = 8.60$. • Black's modified gain coefficient (1.57) exceeded the pre-specified effectiveness criterion of 1.20. • Teachable Machine functioned as a supervised classification tool within a teacher-guided sequence, not as an adaptive tutoring system.	This study examined the effectiveness of a Teachable Machine-based instructional program for developing electronic concepts among eighth-grade students in the Palestinian technology curriculum. A quasi-experimental, nonequivalent pretest-posttest control-group design compared two pre-existing classes, since students were not individually randomized. The sample comprised 80 male students from one school, with 40 in each group. The experimental group completed 12 lessons over six weeks using a sequence that integrated Google Teachable Machine for supervised image classification with PictoBlox for linking classified images to concept names and explanations; the control group studied the same unit conventionally. Achievement was measured with a 25-item Electronic Concepts Test. The experimental group achieved a higher post-test mean ($M = 24.20$, $SD = 0.88$) than the control group ($M = 7.75$, $SD = 2.56$), $t(78) = 38.44$, $p < .001$, 95% CI [15.60, 17.31], Cohen's $d = 8.60$, and Black's modified gain coefficient was 1.57. A carefully structured visual-classification sequence can support immediate acquisition of electronic concepts, but the exceptionally large effect, the single-school sample, and the absence of delayed measurement require cautious interpretation and independent replication.

Article Info: Research Article

Keywords: *artificial intelligence in education, Teachable Machine, electronic concepts, quasi-experimental design, eighth grade*

1. Introduction

Artificial intelligence (AI) is increasingly present in school education, both as content that students need to understand and as a set of tools that may support learning in established subject areas. Contemporary K-12 research has therefore shifted from broad claims about AI toward the design of age-appropriate activities in which learners collect examples, train simple models, test outputs, and reflect on classification processes. Visual, no-code tools are particularly relevant because they permit students to engage with basic machine-learning workflows without advanced programming skills.

* Corresponding author. University College of Science and Technology, Palestine.

e-mail address: mbarghot@gmail.com

Doi: <http://doi.org/10.31681/jetol.1880084>

Received 24 Feb 2026; Accepted 30 Aug 2026

ISSN: 2618-6586. This is an open Access article under the CC BY license.



Google Teachable Machine is a web-based graphical tool for creating supervised classification models from images, sounds, or poses. It is not an intelligent tutoring system and does not independently personalize instruction, diagnose misconceptions, or generate adaptive learning pathways. Its educational value lies in making the data-training-testing cycle visible and usable: learners can assemble labeled examples, train a classifier, test new inputs, inspect output probabilities, and revise the examples when classifications are inaccurate (Carney et al., 2020).

Research on K-12 AI education supports the feasibility of such tools while emphasizing the importance of pedagogy. Visual machine-learning environments mainly support supervised image-recognition tasks and the basic stages of model development (Gresse von Wangenheim et al., 2021). Recent reviews identify Teachable Machine as an age-appropriate tool that is commonly used through constructivist, project-based, and human-computer collaborative approaches (Yim & Su, 2024). Martins et al. (2023) further showed that middle- and high-school students could build an image classifier and understand basic machine-learning concepts through a structured course.

The present study applies this type of tool to disciplinary concept learning rather than to AI literacy alone. The Grade 8 Palestinian technology curriculum includes a unit on sensing and remote control that introduces electronic components, symbols, tools, and functional relationships that may be unfamiliar to students. Based on the researchers' classroom experience, observations, and informal discussions with students and technology teachers, some learners had difficulty distinguishing visually similar components and associating each component with its name, characteristics, and function.

A classification-based instructional sequence may be useful in this context because electronic concepts often depend on recognizing defining visual attributes and differentiating examples from non-examples. However, the effect should be attributed to the instructional design surrounding the tool, not to an unsupported claim that the platform itself is adaptive or intelligent in a tutoring sense. Accordingly, this study investigates whether Teachable Machine, integrated with PictoBlox and explicit teacher guidance, can improve students' acquisition of curriculum-aligned electronic concepts.

2. Literature

2.1. Theoretical Framework

The study is grounded in the cognitive theory of multimedia learning and concept-attainment principles. Multimedia learning theory proposes that learners construct knowledge by selecting relevant verbal and pictorial information, organizing it into coherent representations, and integrating those representations with prior knowledge (Mayer, 2021). The theory does not assume that adding images automatically improves learning; visual and verbal materials must be aligned with the objective and presented without unnecessary processing demands.

Concept attainment involves comparing examples, identifying common attributes, discriminating examples from non-examples, and forming a stable category representation (Bruner et al., 1956). The intervention operationalized these processes through labeled image sets of electronic components, repeated classification trials, discussion of similarities and differences, and presentation of the associated concept names and explanations.

The conceptual model guiding the study was as follows: labeled image examples, model training, classification trials, and concept explanations were expected to direct attention to defining features and support comparison, categorization, retrieval, and verbal-pictorial integration; these cognitive processes were then expected to improve performance on the Electronic Concepts Test. Motivation, planning, decision-making, personalization, and adaptive learning were not measured and are therefore not treated as empirical outcomes.

Table 1. Conceptual model of the instructional mechanism

Instructional features	Proposed cognitive processes	Measured outcome
Labeled image examples; supervised model training; classification trials; concept labels and explanations	Attention to defining features; comparison; categorization; retrieval; integration of pictorial and verbal representations	Acquisition of electronic concepts on the 25-item test

2.2. Related Literature and Research Gap

Carney et al. (2020) presented Teachable Machine as an approachable interface through which users can create custom classification models without specialized technical expertise. Toivonen et al. (2020) reported that primary-school children could use the platform in multidisciplinary co-design activities. Gresse von Wangenheim et al. (2021), in a ten-year systematic mapping, found that K-12 visual machine-learning tools mainly support image-recognition tasks, basic supervised-learning workflows, and deployment through block-based programming environments.

Systematic reviews also reveal important evidence gaps. Yue et al. (2022) found that K-12 AI programs frequently reported positive motivational and attitudinal outcomes, but objective measurement of knowledge acquisition was less common. Casal-Otero et al. (2023) similarly concluded that few initiatives rigorously assessed whether students understood AI concepts following the learning experience. Rizvi et al. (2023) reported generally positive cognitive and affective outcomes but called for context-sensitive pedagogy, stronger empirical designs, and consistent outcome constructs.

More recent evidence positions Teachable Machine as a practical classroom tool rather than an autonomous tutor. Martins et al. (2023) guided 108 middle- and high-school students through an introductory course that included building an image-classification model and found that students could understand basic machine-learning concepts. Yim and Su (2024) identified Teachable Machine among age-appropriate K-12 AI-learning tools and emphasized constructivist and project-based pedagogies.

Most prior work has focused on learning about AI and machine learning. Less evidence is available on using a no-code image classifier as a representational and categorization tool for acquiring concepts within a formal technology curriculum. The present study addresses this applied gap by evaluating a curriculum-specific instructional program with an immediate concepts test and a nonequivalent control-group comparison.

The purpose of this study was to design and evaluate a Teachable Machine-based instructional program for developing electronic concepts in the Grade 8 sensing and remote-control unit. The study addressed the following research questions:

RQ1. What electronic concepts were identified in the target unit, and how were they represented in the instructional program?

RQ2. Were the experimental and control groups comparable on the pretest of electronic concepts?

RQ3. Did the experimental group outperform the control group on the post-test of electronic concepts?

RQ4. Did the experimental group achieve the instructional-effectiveness criterion defined as Black's modified gain coefficient of at least 1.20?

The hypotheses were aligned directly with the quantitative research questions:

H1. The experimental and control groups will not differ significantly on the pretest of electronic concepts.

H2. The experimental group will obtain a significantly higher post-test mean than the control group.

H3. The experimental group will achieve Black's modified gain criterion of at least 1.20.

3. Methodology

3.1. Research Model/Design

A quasi-experimental, nonequivalent pretest-posttest control-group design was used. The design is quasi-experimental because two intact, pre-existing classes were assigned to conditions and individual students were not randomly allocated. Both groups completed the Electronic Concepts Test before and after instruction. The experimental group studied through the Teachable Machine-based program, whereas the control group received the conventional instruction used at the school.

Table 2. Quasi-experimental design

Group	Pretest	Instruction	Post-test
Experimental	Electronic Concepts Test	Teachable Machine-PictoBlox program	Electronic Concepts Test
Control	Electronic Concepts Test	Conventional instruction	Electronic Concepts Test

3.2. Sampling or Study Group

The study was conducted at Saad bin Abi Waqqas Basic School (B) in the North Gaza Education Directorate during the first semester of the 2021-2022 academic year. The sample consisted of 80 male eighth-grade students aged approximately 13-14 years. Class 8/4 served as the experimental group and class 8/6 as the control group, with 40 students in each class. The classes came from the same school and broadly similar educational and socioeconomic context. An additional exploratory sample of 26 students was used during program and instrument development.

3.3. Research Procedures

The instructional program was developed using the analysis, design, development, and evaluation stages of Khamis's instructional-design model. Content analysis of the sensing and remote-control unit identified three major curricular tasks and 25 electronic concepts. The program linked Teachable Machine image categories to PictoBlox displays containing the concept name and explanatory information.

Table 3. Core components of the experimental instructional sequence

Phase	Instructional activity	Learning function
Orientation and concept introduction	Introduce the unit, concept names, prior examples, and classification task.	Activate prior knowledge and establish category labels.
Data preparation and model training	Organize labeled images of electronic components and train the Teachable Machine image classifier.	Attend to defining features and compare examples.
Testing and discrimination	Test unfamiliar images, inspect classification output, identify errors, and refine the examples.	Differentiate categories and examine ambiguous cases.
Concept connection through PictoBlox	Link the classifier output to the component name, definition, and curricular explanation.	Integrate visual and verbal representations.
Guided practice and assessment	Complete teacher-guided comparisons, exercises, and formative checks across 12 lessons.	Consolidate retrieval and application.

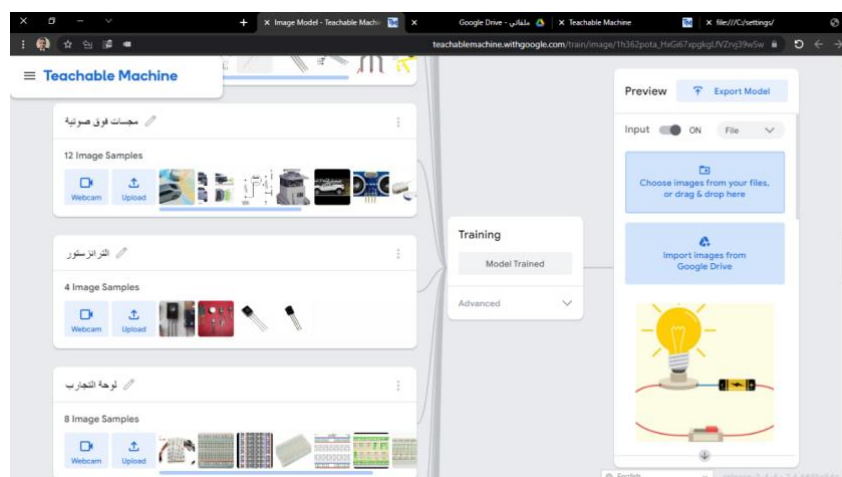


Fig. 1. Example of image categories used to train the Teachable Machine classifier

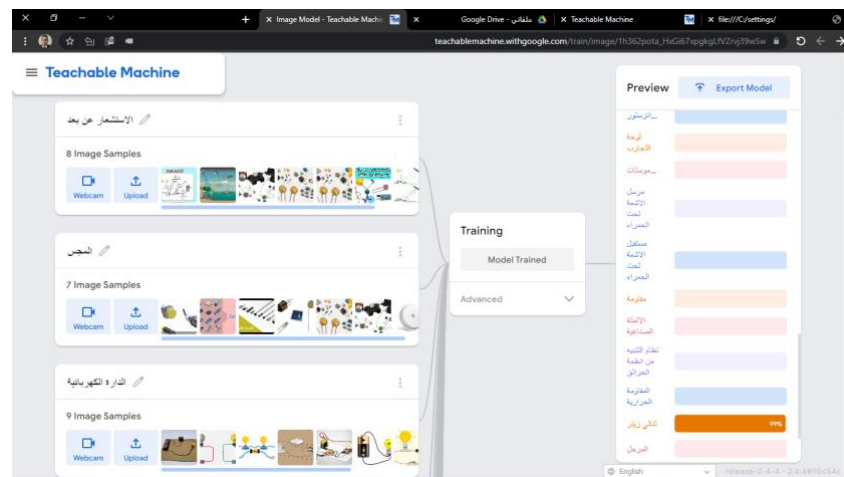


Fig. 2. Example of the trained classification interface used in the instructional program

The control group studied the same curricular unit through the conventional classroom approach followed at the school. The curricular objectives, unit content, lesson period, and testing schedule were maintained; the planned instructional difference was the experimental group's use of the Teachable Machine-PictoBlox sequence.

The pretest was administered on September 1, 2021. Instruction began on September 5 and ended on October 20, 2021. The unit was taught in 12 lessons, at a rate of two lessons per week. In the experimental condition, students viewed and compared labeled examples, participated in training the image classifier, tested new images, inspected the classification output, and connected each output to the concept name and explanation displayed through PictoBlox. The post-test was administered to both groups on October 23, 2021.

The intervention was implemented within regular school instruction in coordination with the school administration and the North Gaza Education Directorate. Only coded achievement scores were used for analysis, no identifying student information was included in the dataset reported in the manuscript, and findings are presented in aggregate form.

3.4. Data Collecting Tools

The Electronic Concepts Test was a researcher-developed, 25-item multiple-choice test aligned with the target unit. Each item had four response options, one correct answer, and a score of 1 for a correct response and 0 for an incorrect response; total scores therefore ranged from 0 to 25. A table of specifications distributed the items across the three-unit topics and six cognitive levels, as shown in Table 4.

Table 4. Specification table for the Electronic Concepts Test

No.	Unit topic	Remembering	Understanding	Application	Analysis	Synthesis	Evaluation	Items	Weight
1	Sensing at a distance	2	1	1	1	0	0	5	20%
2	Making sensing systems	4	2	2	2	1	1	12	48%
3	Sensors everywhere	2	2	1	1	1	1	8	32%
Total	Total	8	5	4	4	2	2	25	100%
Weight	Cognitive-level weight	32%	20%	16%	16%	8%	8%	100%	100%

3.5. Validity and Reliability

Content validity was examined by specialists in curricula, teaching methods, educational technology, and school technology instruction. For item analysis, the 26-student exploratory sample was divided into two equal groups representing the higher- and lower-scoring halves ($n = 13$ each). Difficulty indices ranged

from .40 to .73 and discrimination indices from .37 to .72; the item-level results are presented in Table 5. Item-total correlations ranged from .344 to .683 and were statistically significant at $p < .05$, as reported in Table 6.

Table 5. Item difficulty and discrimination indices

Item	Difficulty	Discrimination	Item	Difficulty	Discrimination
1	0.56	0.53	14	0.40	0.48
2	0.49	0.60	15	0.58	0.50
3	0.53	0.37	16	0.50	0.48
4	0.59	0.65	17	0.68	0.45
5	0.67	0.53	18	0.63	0.47
6	0.72	0.44	19	0.52	0.60
7	0.65	0.51	20	0.57	0.46
8	0.53	0.62	21	0.49	0.61
9	0.49	0.71	22	0.53	0.37
10	0.40	0.57	23	0.67	0.53
11	0.55	0.60	24	0.73	0.45
12	0.69	0.51	25	0.66	0.53
13	0.47	0.48	—	—	—
Overall	0.572	0.522	—	—	—

Table 6. Item-total correlations for the Electronic Concepts Test

Item	Item-total r	p	Item	Item-total r	p
1	0.538	< .001	14	0.555	< .001
2	0.672	< .001	15	0.621	< .001
3	0.593	< .001	16	0.593	< .001
4	0.544	< .001	17	0.544	< .001
5	0.369	.018	18	0.369	.018
6	0.344	.028	19	0.344	.028
7	0.416	.007	20	0.416	< .001
8	0.463	< .001	21	0.463	< .001
9	0.601	< .001	22	0.683	< .001
10	0.669	< .001	23	0.628	< .001
11	0.546	< .001	24	0.538	< .001
12	0.612	< .001	25	0.672	< .001
13	0.418	< .001	—	—	—

Reliability was estimated in two ways. The correlation between the odd- and even-item halves was .663; because the halves were unequal in length, the coefficient was corrected using the Guttman split-half formula, yielding .798. Cronbach's alpha was .825. The reliability evidence is summarized in Table 7 and supports the use of the test for group-level research.

Table 7. Reliability estimates for the Electronic Concepts Test

Reliability estimate	Coefficient	Interpretation
Odd-even half correlation	.663	Pre-correction association between the two test halves
Guttman split-half coefficient	.798	Adequate split-half reliability
Cronbach's alpha	.825	Adequate internal consistency

3.6. Data Analysis

Descriptive statistics were calculated for all test administrations. Independent-samples *t* tests were used to examine pretest comparability and the post-test difference because the two classes were independent groups of equal size. Statistical significance was evaluated at $\alpha = .05$, and exact results are reported as $p < .001$ when appropriate. Cohen's *d* was calculated using the pooled standard deviation. Black's modified gain coefficient was calculated for the experimental group using the pretest mean, post-test mean, and maximum test score. The post-test *t* statistic, confidence interval, and effect size were calculated from the group means, standard deviations, and sample sizes presented in the study tables.

4. Findings and Discussions

4.1. Identified Concepts and Program Representation

Content analysis produced a validated set of 25 electronic concepts distributed across the three topics of the sensing and remote-control unit. These concepts informed the test blueprint, the labeled image categories, and the PictoBlox explanations. The program was therefore aligned at three levels: curriculum content, instructional representation, and assessment.

4.2. Pretest Comparability

The experimental group obtained $M = 8.60$ ($SD = 3.73$), and the control group obtained $M = 7.83$ ($SD = 2.29$). The difference was not statistically significant, $t(78) = 1.12$, $p = .266$, indicating no evidence of a baseline difference on the measured electronic concepts.

Table 8. Pretest comparison between the experimental and control groups

Group	n	Pretest M	SD	t(78)	p
Experimental	40	8.600	3.734	1.12	.266
Control	40	7.825	2.285		

4.3. Post-test Difference

The experimental group achieved a post-test mean of 24.20 ($SD = 0.88$), compared with 7.75 ($SD = 2.56$) for the control group. An independent-samples t test yielded $t(78) = 38.44$, $p < .001$, with a mean difference of 16.45 points and a 95% confidence interval from 15.60 to 17.31. Cohen's d was 8.60. Accordingly, H2 was supported.

Table 9. Post-test comparison and effect size

Group	n	Post-test M	SD	t(78)	Sig.	95% CI
Experimental	40	24.200	0.880	38.44	< .001	[15.60, 17.31]
Control	40	7.747	2.560			

4.4. Black's Modified Gain

Using an experimental-group pretest mean of 8.60, a post-test mean of 24.20, and a maximum possible score of 25, Black's modified gain coefficient was 1.57. This exceeded the prespecified criterion of 1.20; therefore, H3 was supported.

Table 10. Black's modified gain coefficient for the experimental group

Group	Pretest M	Post-test M	Maximum score	Black modified gain	Criterion
Experimental	8.60	24.20	25	1.57	≥ 1.20

4.5. Discussion

The results show a substantial immediate post-test advantage for students who completed the Teachable Machine-PictoBlox instructional sequence. The finding is consistent with the proposed mechanism: repeated exposure to labeled visual examples, explicit comparison of component features, classification trials, and immediate presentation of the concept label may have helped students organize and retrieve the defining attributes of the electronic concepts.

The result is theoretically compatible with multimedia learning and concept-attainment perspectives. The program combined pictorial examples with verbal labels and explanations and required learners to discriminate among categories. The evidence does not demonstrate that Teachable Machine personalized instruction, dynamically adapted a learning pathway, or functioned as an intelligent tutor. It was a supervised classification tool embedded within a teacher-designed sequence.

The result also extends prior work showing that students can use accessible visual platforms to build and test classifiers (Carney et al., 2020; Martins et al., 2023; Toivonen et al., 2020). Whereas much of that literature concerns learning about machine learning, this study used the workflow to support concepts in a

formal technology unit. The contribution is consequently pedagogical and application-based rather than a novel AI architecture.

The magnitude of the observed difference is unusual for classroom research. The experimental mean was close to the test maximum, and the reported score distributions show little overlap. This pattern may reflect intensive alignment between instruction and assessment, a ceiling effect, the specificity of the test, or an exceptionally strong short-term treatment effect. The result should therefore be interpreted as preliminary and bounded by the available summary statistics. Replication with raw-data auditing, distributional diagnostics, additional schools, and delayed outcomes is necessary.

5. Conclusion and Suggestions

A structured program combining Teachable Machine image classification, PictoBlox concept presentation, and teacher-guided comparison was associated with higher immediate achievement in electronic concepts among the participating eighth-grade students. The study provides preliminary evidence that accessible machine-learning tools may support curriculum learning when they are embedded in an explicit pedagogical design. The findings should not be interpreted as evidence that Teachable Machine is an adaptive tutor or that AI tools are uniformly effective across contexts. The quasi-experimental design, single-school sample, immediate outcome, and exceptionally large effect require bounded interpretation and replication.

5.1. Limitations

The study involved two intact classes from one boys' school; student-level random assignment did not occur. The outcome was measured with one researcher-developed immediate achievement test, without delayed retention or transfer measures. The near-ceiling experimental mean and extremely large standardized effect restrict the certainty and generalizability of the estimate. The study did not directly measure motivation, planning, decision-making, adaptive learning, or broader AI literacy. Implementation fidelity and teacher effects were not measured independently.

5.2. Suggestions

Teachable Machine should be described and used as a supervised classification and visualization tool, not as an adaptive or personalized tutoring system. Image categories, concept definitions, classroom explanations, and assessment specifications should be aligned before implementation. Teachers should provide multiple examples and non-examples, discuss classification errors, and distinguish model output from scientifically verified knowledge. Future studies should use larger, mixed-gender, multi-school samples and include delayed retention, transfer, engagement, implementation-fidelity, and qualitative process measures. When intact classes are used, future analyses should consider ANCOVA or multilevel models where the sample and number of clusters permit.

Declarations

Ethics approval and consent to participate: The intervention was conducted within regular classroom instruction with administrative coordination from the school and the North Gaza Education Directorate. Only coded, non-identifiable achievement data were analyzed and reported in aggregate.

Consent for publication: The manuscript contains no personally identifiable participant information.

Availability of data and materials: De-identified data and the instructional materials are available from the corresponding author upon reasonable academic request, subject to applicable school and institutional requirements.

Competing interests: The authors declare that they have no competing interests.

Funding: The authors report that the study received no external funding.

Author contributions: Mahmoud M. F. Barghot contributed to conceptualization, methodology, instrument development, statistical analysis, and preparation of the original manuscript. Ahmed Mohamed

Abu Elba contributed to instructional-program development, classroom implementation, data collection, and critical review of the manuscript. Both authors reviewed and approved the final version.

References

- Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. Wiley.
- Carney, M., Webster, B., Alvarado, I., Phillips, K., Howell, N., Griffith, J., Jongejan, J., Pitaru, A., & Chen, A. (2020). Teachable Machine: Approachable web-based tool for exploring machine learning classification. In *Extended abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–8). Association for Computing Machinery.
- Casal-Otero, L., Catalá, A., Fernández-Morante, C., Taboada, M., Cebreiro, B., & Barro, S. (2023). AI literacy in K–12: A systematic literature review. *International Journal of STEM Education*, 10, Article 29.
- Gresse von Wangenheim, C., Hauck, J. C. R., Pacheco, F. S., & Bueno, M. F. B. (2021). Visual tools for teaching machine learning in K–12: A ten-year systematic mapping. *Education and Information Technologies*, 26, 5733–5778.
- Khamis, M. A. (2013). *Educational computer and multimedia technology* [In Arabic]. Dar Al-Sahab.
- Martins, R. M., Gresse von Wangenheim, C., Rauber, M. F., & Hauck, J. C. R. (2023). Machine learning for all! Introducing machine learning in middle and high school. *International Journal of Artificial Intelligence in Education*. Advance online publication.
- Mayer, R. E. (2021). *Multimedia learning* (3rd ed.). Cambridge University Press.
- Rizvi, S., Waite, J., & Sentance, S. (2023). Artificial intelligence teaching and learning in K–12 from 2019 to 2022: A systematic literature review. *Computers and Education: Artificial Intelligence*, 4, Article 100145.
- Toivonen, T., Jormanainen, I., Kahila, J., Tedre, M., Valtonen, T., & Vartiainen, H. (2020). Co-designing machine learning apps in K–12 with primary school children. In *2020 IEEE 20th International Conference on Advanced Learning Technologies* (pp. 308–310). IEEE.
- Yim, I. H. Y., & Su, J. (2024). Artificial intelligence learning tools in K–12 education: A scoping review. *Journal of Computers in Education*. Advance online publication.
- Yue, M., Jong, M. S. Y., & Dai, Y. (2022). Pedagogical design of K–12 artificial intelligence education: A systematic review. *Sustainability*, 14, Article 15620.