Humanities & Social Sciences Communications

REVIEW ARTICLE

https://doi.org/10.1057/s41599-023-02338-x OPEN

A meta-analysis to gauge the impact of pedagogies employed in mixed-ability high school biology classrooms

Malavika E. Santhosh[1], Jolly Bhadra[1✉], Zubair Ahmad[1] & Noora Al-Thani[1]

This article systematically reviews the pedagogies employed in mixed-ability high school biology classrooms to spotlight the most effective educational model, in terms of learning gains. A meta-analysis was performed on 32 eligible studies sorted via the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology. The effect sizes (impacts) were calculated in terms of students' affective, behavioral, and cognitive gains. The results confirmed that the overall effectiveness of non-traditional models was highly significant when compared to traditional lecture models ($p < 0.001$). Furthermore, this impact is even more profound when problem-based, inquiry-based, and argumentation-based approaches are adopted, contributing to students' cognitive, affective, and behavioral gains. Further findings also propound the necessity for future studies evaluating affective gains during project-based, problem-based, and argumentation-based models and behavioral gains during blended models. This study would benefit researchers, policymakers, and academicians to innovate and implement novel pedagogical strategies, considering the students' learning gains in mixed-ability biology classrooms.

Introduction

Biology is a graphic science, wherein the learners study living organisms and their interactions with one another and their environments (Goodwin and Dawkins, 1995, p. 47). This is a broad definition since the scope of biology is vast. High school biology students often feel difficulty in mastering the subject. The potential reason for this is due to complex materials, invisible/intangible objects, and complex terms (Cimer, 2012). This difficulty in learning could influence students' variables such as their performance, attitudes, behaviors, knowledge, skills, etc. Furthermore, traditional education has not been efficacious in solving these issues (Ebrahim and Naji, 2021; Yapici and Akbayin, 2012). Thus, many non-traditional pedagogical models/approaches are being developed, tested, and implemented for the efficient teaching of biology concepts. The reported non-traditional, students centered pedagogical models in high school biology classrooms include inquiry-based, problem-based, project-based, virtual simulation-based, game-based, argument-based, etc. (Klisch et al., 2013; Nunaki et al., 2019; Ping et al., 2020; Sivia et al., 2019a; Thisgaard and Makransky, 2017; Thurrodliyah et al.,

[1] Qatar University Young Scientists Center, Qatar University, Doha 2713, Qatar. ✉email: jollybhadra@qu.edu.qa

2020; Yapici and Akbayin, 2012). Furthermore, a combination of these models has also been reported (Anazifa and Djukri, 2017; Lui and Slotta, 2014; Thompson et al., 2020).

Despite the various pedagogies employed, it is often tricky to conclude the best pedagogical approach in high school biology education. Even the literature shows disproportionate results with varying levels of effectiveness. Such as the reported impact of problem-based learning is [$d = 0.89$ (Xu et al., 2021)], project-based learning are [$d = 1.36$ (Balemen and Keskin, 2018; $d = 0.95$ (Ayaz and Söylemez, 2015)], inquiry-based are [$d = 1.26$ (Funa and Prudente, 2021); $d = 0.35$ (Wang et al., 2011)] and web-based is [$d = 0.668$ (Bayraktar, 2001)] in biology education. This difference in effect sizes might be due to the diversity in students (low/average/high/mixed achievers, gifted, at-risk students, etc.) and their pedagogical requirements (cognitive, affective, behavioral learning gains). Different students may benefit disproportionately from different learning gains (Korkor Sam et al., 2018; Steenbergen-Hu et al., 2020).

Therefore, this article seeks to compute the impact of various non-traditional educational models, particularly for a mixed-ability (low/average/high achievers) biology classroom in terms of students' learning gains (cognitive, affective, and, behavioral). This meta-analysis intends to compute the impact of non-traditional (modern) pedagogical models and offer pedagogical justifications for the requirement of efficacious and appropriate practices in high school biology. For this, the study has first investigated the diverse pedagogical approaches that are being employed in high school biology classrooms. The overall effectiveness of the non-traditional pedagogies compared with the traditional lecture model has also been inspected. Subsequently, a comparative examination of different teaching approaches has been conducted, taking into account the different learning gains. Therefore, the research questions addressed in the paper are:

1. What are the various non-traditional pedagogical models/approaches employed in high school biology education?
2. What is the overall impact of the non-traditional pedagogical models employed for mixed-ability high school biology classrooms (when compared to the traditional lecture model)?
3. What is the comparative effectiveness of each pedagogical model, concerning the students' gains (cognitive, affective, behavior)?

Review of literature and conceptual framework

non-traditional pedagogies in biology education. Literature has witnessed the implication of diverse teaching approaches that are often employed explicitly (i.e., targeting a single pedagogical model) or in conjugation (targeting two/more models). Dichotomizing these, project-based pedagogy in biology, a student-centered and multidisciplinary approach, is often employed, where students work on projects to investigate and find answers to complex questions/problems. Prior literature reports that project-based models are often employed for teaching genetics (Sivia et al., 2019a) and animal physiology such as digestive/circulatory/respiratory systems (Sukmawati et al., 2019; Anazifa and Djukri, 2017; Blacer-Bacolod, 2022). This approach in high school biology has reported improved knowledge, critical thinking skills, civic engagement, conceptual understanding, and application skills (Sivia et al., 2019a; Sukmawati et al., 2019; Anazifa and Djukri, 2017; Blacer-Bacolod, 2022; Sukmawati et al., 2019; Sari et al., 2019). In contrast, the problem-based model is often single subject/problem-based, employing case studies/fictitious scenarios as the problem, which helps students acquire 21st-century quintessential skills of problem-solving. Problem-based model is often used to instruct and educate students on environmental issues (Özalemdar, 2021; Thurrodliyah et al., 2020; Thinkhamchoet et al., 2021). Many studies have already shown enhanced students' skills of reflection, knowledge, creativity, critical thinking, positive attitudes and behavior, and psychomotor learning outcomes (Anazifa and Djukri, 2017; Thurrodliyah et al., 2020; Özalemdar, 2021; Hugerat et al., 2021; Kolarova et al., 2014). The primary similarity between the project-based and problem-based models is their focus on open-ended questions/problems, driving the inquiry process. Another model is the inquiry-based pedagogical model that supports students to acquire knowledge independently via the inquiry process (Hadjichambis et al., 2022; Nunaki et al., 2019; Wilson et al., 2010). Studies have reported the successful execution of inquiry-based models for environmental biology and biodiversity (Ristanto et al., 2022; Lui and Slotta, 2014; Hadjichambis et al., 2022), and cell biology (Saputri et al., 2019; Thompson et al., 2020; Ping et al., 2020). Many studies have reported better students' achievements, metacognitive skills, perception, and, the conception of biology learning via the inquiry approach (Hadjichambis et al., 2022; Kagnici and Sadi, 2021; Nunaki et al., 2019; Ristanto et al., 2022; Saputri et al, 2019; Wilson et al., 2010). Inquiry-based teaching has also been executed in collaboration with argument-based models (Ping et al., 2020; Ristanto et al., 2022), and game-based models (Lui and Slotta, 2014; Thompson et al., 2020) for high school biology. And, it is reported to promote critical thinking, argumentation skills, and science process skills, (Lui and Slotta, 2014; Ping et al., 2020; Ristanto et al., 2022; Thompson et al., 2020). Therefore, other than typical pedagogical alternatives (e.g., project-based, problem-based, inquiry-based, etc.); the gamification approach has gained much traction. A possible justification is the contribution of the game-based model toward students' engagement and interactivity, which are supposed to be the most challenging variables that should be considered (Cai et al., 2022). Similarly, modeling and virtual simulations have also been reported to be impactful in promoting the knowledge and understanding of complicated biological concepts such as cell biology and genetics (Marbach-Ad et al., 2008; Mulder et al., 2016; Thisgaard and Makransky, 2017; Li and Ma, 2010). Also, a careful blend of pedagogical approaches (termed as the blended model) in a wisely framed online and offline setting is often recommended to improve students' cognitive skills, achievements, and attitudes toward the course contents and the internet (Yapici and Akbayin, 2012; Ebrahim and Naji, 2021; Kazu and Demirkol, 2014).

Students' learning gains. A systematic review of diverse high school pedagogies in association with learning gains seems to be an underexplored area (in the context of biology). The notion of students' learning gain is defined as growth or change in knowledge, skills, and attitudes over time (Cronbach and Furby, 1970; Roohr et al., 2017). The affective learning gains account for attitudes, confidence, motivation, satisfaction, and well-being. While behavioral learning gains account for students' behavioral skills such as engagement, leadership skills, and teamwork. Cognitive learning gains pertain to skills associated with cognitive growth, including comprehension, information retention, critical thinking, creative thinking, logical reasoning, analytical thinking, and scientific reasoning. (Bloom et al., 1956). In psychology and education, the interlinked affective, behavioral, and cognitive learning gains were used to understand and unravel the multidimensional notions of learning gains (Ostrom, 1969). However, the range of learning gains evaluated in the literature seems to be extensively diverse (in the context of biology education). For example, a longitudinal study on nearly 17000 students across 50 US colleges was conducted to study a range of learning gains. Their study reported positive learning gains in engagement, critical thinking, moral reasoning, and leadership; while negligible learning gains in literacy, and political and social engagement

(Pascarella and Blaich, 2013). In contrast, some studies evaluated only particular kinds of learning gains. For instance, some studies only concentrated on the cognitive learning gains of biology students (Nunaki et al., 2019; Özalemdar, 2021; Ping et al., 2020; Ristanto et al., 2022; Thompson et al., 2020). While some others focused only on affective gains (Anazifa and Djukri, 2017; Brom et al., 2011; Venville and Dawson, 2010). This diversity in evaluating the learning gains might be predisposed by the diversity in student needs. This was well explained by a meta-analysis and systematic review by Steenbergen-Hu et al., 2020). Their study has reported that gifted underachievers benefit more from pedagogies focusing on affective and behavioral gains (such as self-efficacy, goal evaluation, positive perceptions, motivation, and psychosocial functioning), rather than cognitive gains such as course grades (Steenbergen-Hu et al., 2020). In this context, Korkor Sam et al. (2018) insist on the 3E learning model (exploration, explanations, and expansion) for improving the performance of low achievers in biology (Korkor Sam et al., 2018). Recommendations by Chaplin (2007), include modeling and coaching learning in classrooms to augment the critical thinking skills of students at risk in introductory biology courses. Interestingly a study by Yaduvanshi and Singh, 2019 has recommended structured cooperative learning for mixed achievers (low, average, and high achievers) in secondary biology classrooms (Yaduvanshi and Singh, 2019). Therefore, concluding the best pedagogical model in a mixed-ability high school biology must contemplate the wide-ranging students' needs (cognitive, affective, and behavioral gains in collaboration) (Wei et al., 2021).

Methods

The research employed a systematic literature review approach in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. (Page et al., 2021). The screening for studies as per PRISMA methodology was employed in three stages (Fig. 1). The three stages include (1) Identifying the articles employing a specific search strategy; (2) Initial screening of the articles based on title, abstract, and content relatedness; and (3) Final screening of the studies based on predefined inclusion/exclusion criteria.

Search strategy. The relevant research articles were gathered using widely used web search engines like "Web of Science," "Journals for Educational Research Information Center (ERIC)," and "Scopus". Specific keywords were included for searching the articles: ("STEM" OR "science" OR "biology" OR "genetics" OR "anatomy" OR "botany" OR "zoology") AND ("teach*" OR "learn*" OR "pedagog*" "education" OR "*school" OR "grade 10" OR "grade 11" OR "grade 12") AND ("online" OR "digital" OR "laboratory" OR "project" OR "discovery" OR "blended" OR "flip*" OR "inquiry" OR "discovery" OR "problem-*" OR "game" OR "virtual-*" OR "immersive"). All the papers included in the study were empirical research studies.

A total of 799 papers were located through keyword searches: Scopus ($n = 330$), ERIC ($n = 212$), and Web of Science ($n = 257$). The identified records were exported to Excel to find the duplicates. Wherein, 327 articles were discarded due to repetition. The articles were screened for apt title, gist, abstract, appropriate context, inclusion/exclusion criteria, retrieval, and relevancy (Fig. 1). The initial selection of the studies was based on skim-reading the title and abstract. Followingly articles have been thoroughly investigated for context-relatedness. And finally, the whole article has been reviewed for relevance. The context-relatedness and relevancy were checked via inclusion and exclusion criteria, set under the scope of the review paper.

Inclusion/exclusion criteria. The inclusion criteria for article selection were: Biology education model at senior high school (e.g., Grades 10–12); Articles extracted from 2008–2022 March; Peer-reviewed articles; Articles in English. The exclusion criteria were: Education level other than high school (e.g., junior high school, middle school, UG, PG, etc.); The teaching models for non-biology subjects (chemistry, physics, language, mathematics, etc); and qualitative research and review papers. The final exclusion criteria were as follows: non-empirical studies; studies lacking a pre-/post-test or control/experiment design; studies with data that did not conform to the Comprehensive Meta-Analysis (CMA) format (e.g., mean, standard deviation, Cohen's d, t-value, p value, etc.); studies presenting results based on non-student variables (e.g., teachers' perception, author's opinions, etc.); insufficient data for calculating effect size (i.e., p value/t-value $= 0$). Ultimately, 32 peer-reviewed empirical articles met all the criteria and were included in the meta-analysis Table 1 displays the descriptive characteristics of the 32 selected studies considered for the meta-analysis review: (a) studies, (b) pedagogical model (c) sub-pedagogical approach, (d) study design, (e) participants' grade, (f) topic of biology, (g) sample size, (h) study results, and, (i) country of publication.

Figure 2 illustrates the distribution of included studies categorized by continent of publication, year, and grade. It can be noted in Fig. 2a, that more studies were from Asia ($n = 13$), followed by the Middle East ($n = 7$), America ($n = 6$), Europe ($n = 5$), and Australia ($n = 1$). Thus, research studies from almost all continents were included in the study. When analyzing the grades (in Fig. 2b), studies incorporating the grade 10 participants only were the most ($n = 16$), followed by grade 11 ($n = 7$). Figure 2c depicts the studies by the year of publication. The studies were extracted from the past 15 years (2008–March 2022), and most of the studies were extracted from the years 2019 and 2021 ($n = 5$ each).

Coding procedure. The coding of the studies was conducted by two authors, who have more than 15 years of educational research experience. They read each paper individually and employed the content analysis technique to extract the data, (Hsu et al., 2013). Consecutively, an inter-rater-reliability test has been computed to analyze how often two coders agree with each other. Inter-rater reliability using Cohen's kappa statistic was employed and this value was found to be 0.93, corresponding to "almost perfect agreement" (Warrens, 2015). Occasional disagreements between the coders were resolved by discussions and consensus. The quality of the shortlisted studies was evaluated by using the Medical Education Research Study Quality Instrument (MERSQI) (Reed et al., 2007). The instrument comprised 10 constructs, encompassing six domains related to study quality, including study design, sampling, type of data, validity of evaluation instrument, data analysis, and outcomes. The potential total score ranged from 5 to 18, with this research attaining a mean score of 15.61 (SD $= 0.88$). This suggests that the studies incorporated into our meta-analysis demonstrated commendable overall methodological quality (Smith and Learman, 2017).

For the categorization of studies, in terms of pedagogical approach and learning gains, terms and keywords used by the respective authors of the shortlisted studies were considered.

The meta-analysis and interpretation. The meta-analysis was conducted using the CMA software package, specifically version 3.3.070. The Der Simonian and Laird methods were employed to calculate both individual and average effect sizes, along with 95% confidence intervals (DerSimonian and Laird, 2015). This analysis aimed to elucidate the influence of diverse pedagogical approaches on high school students. The raw empirical data were extracted

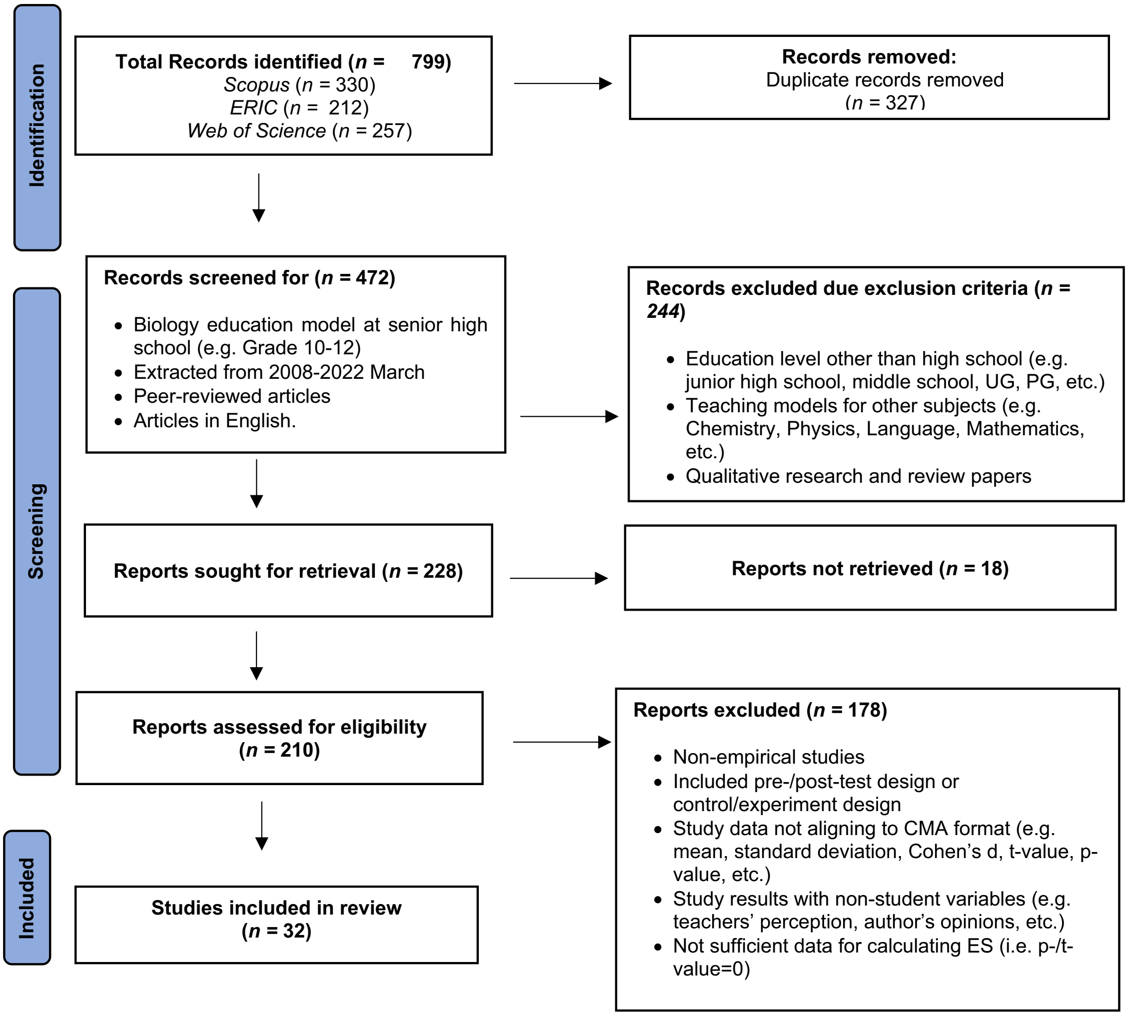

Fig. 1 PRISMA 2020 flow diagram. The figure reveals the inclusion-exclusion criteria for studies included in the meta-analysis.

from the finalized studies in the form of pre-/post- or control/treatment means, SDs, *t*-values, *p* values, etc., to calculate the effect size (Cohen's *d* index) using CMA software. A forest plot diagram revealed the distribution of the effect sizes (at 95 percent confidence intervals) of all the shortlisted studies. As per Cohen's classification, Cohen's *d* values of ≤0.2, approximately 0.5, and ≥0.8 are considered to indicate low, moderate, and high levels of significance, respectively (Cohen, 2013). Furthermore, according to Arnold, any effect size (Cohen's *d*) index=0.4 was considered to be educationally impactful (Arnold, 2011). Most of the studies included in the review were large sample-sized, therefore Cohen's *d* value was preferred over Hedge's *g* value for measuring the effect sizes. A random-effects model was used to compute the mean effect sizes. This model assumes that the variation in the effect sizes of the individual studies is due to sampling error or study design differences. In random-effects analyses, it assumes that each study tends to have a different "true" effect. Thus, the model has been used to account for the heterogeneity between the studies. The impact of each pedagogical model is further investigated by a subgroup meta-analysis and has been explained in the results and discussion section. Before analysis, tests for heterogeneity, test for publication bias, and sensitivity tests were performed.

Test for heterogeneity. The heterogeneity test in meta-analysis determines the variation in study outcomes between studies. To check the heterogeneity of the studies, Cochran's *Q* statistic and I^2 statistic were used (Cochran, 1954). A significant *Q* statistic indicates that the effect sizes are derived from diverse populations, indicating heterogeneity. Conversely, a non-significant *Q* statistic suggests that all studies are presumed to share the same population effect, which implies the application of fixed-effect models.

In our study, a random-effects model was employed because of the assumption that effect sizes vary across studies due to true variance and also due to the fact of reporting the means for the universe of all comparable studies (Borenstein et al., 2009). The difference in effect sizes of the studies could be due to interference of variables (such as sampling error or research design); Therefore, to identify the interfering variable, the heterogeneity analysis was performed using *Q* and I^2 statistics. In our study, the *Q* statistic is also employed as a test for the null hypothesis. When the *Q* value deviates from the degrees of freedom (Df), we reject the null hypothesis, which posits that the true effect size is consistent across all studies. In this instance, the *Q*-value is 401.194 with 31 degrees of freedom. Consequently, we reject the null hypothesis, signifying that true effect size values differ among the studies.

On the other hand, the I^2 statistic represents the proportion of observed variance in effect sizes compared to the actual variance. I^2 values falling within the ranges of <20%, 20–50%, 50%–75%, and ≥75% are interpreted as indicating low, moderate, high, and very high levels of heterogeneity, respectively. Increased I^2 values reveal low dispersion. The study estimates an average I^2 value of 92.273%, showing lower dispersion (Cochran, 1954).

Table 1 Descriptive features of the 32 shortlisted studies for the meta-analysis.

S.No.	Study	Pedagogical model	Sub-pedagogical approach	Study design	Grade	Biological topic	Sample size	Impact of pedagogy	Country
1	(Ristanto et al., 2022)	Inquiry-based & argumentation-based	Guided discovery learning model - argument mapping	Quasi experimental and pre-post-test control group	Grade 10	Environmental changes	89	Enhanced critical thinking skills	Indonesia
2	(Su et al., 2014)	Game-based	Educational card game	Quasi-experimental research	Grade 11	Immunology	72	Enhanced knowledge, learning efficiency & positive perception	Taiwan
3	(Saputri et al., 2019)	Inquiry-based	Stimulating higher-order thinking model (inquiry learning model)	Quasi-experimental research	Grade 12	Cell metabolism	65	Enhanced critical thinking skills	Indonesia
4	(Venville and Dawson, 2010)	Problem & Argumentation-based	Classroom-based argumentation model	Quasi-experimental research	Grade 10	Genetics	92	Better argumentation skills, informal reasoning, and conceptual understanding of the topic	Australia
5	(Ebrahim and Naji, 2021)	Blended model	Flipped learning	Quasi-experimental research.	Grade 10	Cell division	37	Augmented students' attainment of earning	Kuwait
6	(Kolarova et al., 2014)	Problem-solving model	Reflexive model	Pre-post test design	Grade 9–10	Molecular genetics	25	Increased skills of reflection	Bulgaria
7	(Lui and Slotta, 2014)	Game-based & inquiry-based	Immersive simulation and inquiry-based	Design-based research method	Grade 11	Evolution and biodiversity	75	Better engagement and learning	Canada
8	(Sivia et al., 2019a)	Project-based		Mixed-method research	Grade 10	Chemistry, aquaponics, and genetics	30	Enhanced student civic engagement	Columbia
9	(Thompson et al., 2020)	Game-based & inquiry-based	Inquiry-based 3-d virtual reality (VR) game	Pre-post test design	Grade 10–12	Cell Biology	154	Improved students' interactivity and engagement with an enhanced understanding of the concepts	U.s.a.
10	(Özalemdar, 2021)	Problem-solving model	Active learning method	Pre-post-test design	Grade 10	Current environmental issues and human	50	positive effect on the students' environmental attitude and environmental behavior	Turkey
11	(Thisgaard & Makransky, 2017)	Virtual simulation	Virtual laboratory simulation	Mixed method	Grade 12	Biotechnology	128	Increased knowledge, and self-efficiency in biology-related tasks	Denmark
12	(Hugerat et al., 2021)	Problem-solving model	Problem-based—jigsaw discussion	Quasi experimental design	Grade 10	Diseases of the blood circulatory system	204	Improved students' motivation and their perception of the classroom climate.	Israel
13	(Hadjichambis et al., 2022)	Inquiry-based	The Environment Educational Citizenship Model (EEC)	Quasi experimental design	Grade 10	Protection of endangered species	50	Increased students' EC learning gains (EC competencies and EC future actions)	Cyprus
14	(Nunaki et al., 2019)	Inquiry-based	Materials-oriented inquiry-based	Pre-post design	Grade 10	General Biology	70	Increased students' metacognitive skills.	Indonesia
15	(Kagnici and Sadi 2021)	Inquiry-based	5 E (Engagement, Exploration, Explanation, Elaboration, and Evaluation) stem activity-based	True experimental design	Grade 11	Nervous system	99	Better students' academic performance and conceptions of learning Biology	Turkey
16	(Sukmawati et al., 2019)	Project-based	Project-based collaborative learning	Quasi-experimental research	Grade 11	Blood circulatory system	75	Enhanced conceptual understanding or concept application.	Indonesia
17	(Anazifa and Djukri, 2017)	Project-based & problem-based	Project-based and problem-based	Quasi-experimental research	Grade 11	Respiratory system	102	Enhanced student's creativity and critical thinking	Indonesia

Table 1 (continued)

S.No.	Study	Pedagogical model	Sub-pedagogical approach	Study design	Grade	Biological topic	Sample size	Impact of pedagogy	Country
18	(Thurrodliyah et al., 2020)	Problem-solving model	Brain-based model based on socio-scientific issues	ADDIE (Analyze, Design, Develop, Implement, and Evaluate) a research model	Grade 10	Environmental issues	36	Increased students' psychomotor learning outcomes and critical thinking skills	Indonesia
19	(Sari et al., 2019)	Project-based		Mixed method	Grade 11	Research	53	Enhanced students' research skills.	Indonesia
20	(Thinkhamchoet et al., 2021)	Problem-based	Team activity-based learning	Pre-post-test design	Grade 10–12	Environmental protection	100	Improved students' knowledge about environmental conservation and environmental ethics	Thailand
21	(Klisch et al., 2012)	Game-based	Online game-based	Pre-post-test design	Middle and high school	Alcohol abuse	334	Better content knowledge and students' attitudes toward science	The U.S.A.
22	(Mulder et al., 2016)	Virtual simulation	Online scientific modeling	Quasi-experimental design	Grade 10	The human glucose-insulin regulatory system	70	Enhanced students' knowledge and reasoning skills	Netherlands
23	(Wilson et al., 2010)	Inquiry-based	5 e model	Lab-based randomized control study	Grade 10	Sleep disorders and biological rhythms	58	Significantly higher levels of achievement	The U.S.A.
24	(Blacer-Bacolod, 2022)	Project-based	Project-based blended learning by video-making	Mixed method	Grade 10	Animal physiology	92	Improved students' knowledge	Philippine
25	(Yapici and Akbayin, 2012)	Blended model		Pre-post and control-group model	Grade 9–10	Classifications and biodiversity	107	Improved students' biology achievement and attitudes toward the internet	Turkey
26	(Marbach-Ad et al., 2008)	Virtual simulation	Computer animation and illustration activities	Three group study	Grade 11–12	Molecular genetics	248	Improved students' knowledge.	Israel
27	(Ping et al., 2020)	Inquiry-based & argumentation-based	Argumentation-based	Quasi-experimental design	Grade 10	Osmosis and diffusion	112	Significant improvement in argumentation skills, science process skills, and biology understanding	Malaysia
28	(Lham and Sriwattanarothai, 2018)	Game-based	Boardgame	Quasi-experimental design	Grade 10	Cell cycle	25	Enhanced students' conceptual understanding and achievement	Bhutan
29	(Brom et al., 2011)	Game-based	Online micro game	Quasi-experimental design	Grade 10–12	Animal learning	100	Comparable knowledge gains increased overall appeal toward the topic and better retention.	Prague
30	(Lokayut and Srisawasdi, 2014)	Game-based	Computer game-based	Quasi-experimental design	Grade 11	Circulatory system	31	Increased students' perception (motivation)	Thailand
31	(Klisch et al., 2013)	Game-based	Online game-based	Quasi-experimental design	Grade 11–12	Drug abuse	179	Negative students' attitudes towards prescription drug abuse	U.S.A.
32	(Kazu and Demirkol, 2014)	Blended model		Quasi-experimental design	Grade 10–12	Genetics	54	Better academic achievement	Turkey

Test for publication bias. Publication bias exists when the results of an experiment influence the decision of its publication, which is often not recommended in meta-analysis research. To assess the potential presence of publication bias in this study, several methods were employed. These included a funnel plot (Borenstein et al., 2009), a trim-and-fill model (Duval and Tweedie, 2000), and a classic fail-safe N (Rosenthal, 1979). The funnel plot visually represents the association between effect sizes and their corresponding standard errors (SE). A symmetrical funnel plot suggests an absence of publication bias. This study's findings reveal an asymmetrical funnel plot (Fig. 3) corresponding to publication bias due to small-size studies. Therefore, other methods (classic fail-safe N, Trim and fill, and Eggers' linear regression tests) to compute the publication bias were employed. Examination of publication bias employing Egger's test suggested that publication bias would not impact the study analysis and results ($p > 0.05$). Also, the difference between the observed and adjusted estimates was found to be 39.1%, which fell into the

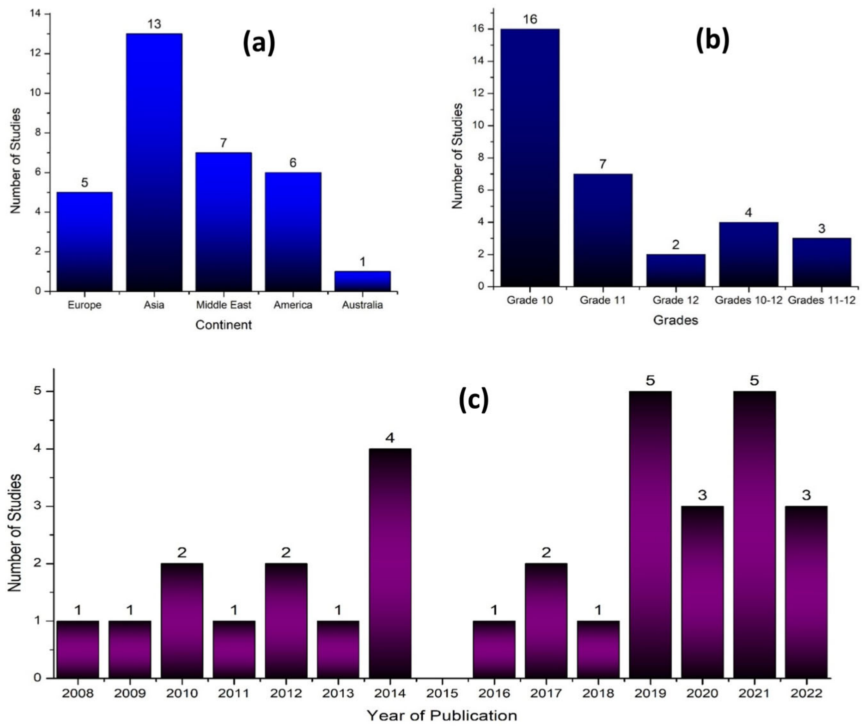

Fig. 2 Features of shortlisted studies. The bar graphs demonstrate the number of studies included in the meta-analysis by **a** continent of publication, **b** by grades, and **c** by year of publications.

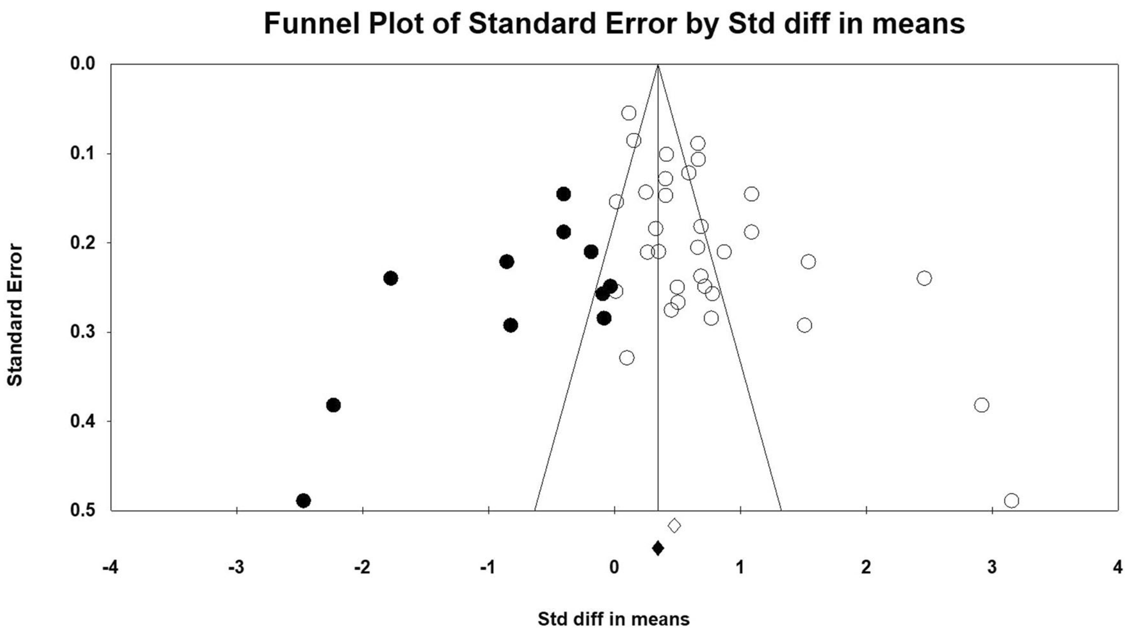

Fig. 3 A funnel plot. The diagram represents the standard error by effect sizes to assess the potential publication bias.

"moderate" cut-off value (20% to 40%), indicating the publication bias to be negligible. The classic fail-safe N was also computed to examine the number of missing studies to prove the non-significance of the meta-analysis conducted due to publication bias (Rosenthal, 1979). The presence of publication bias is not deemed significant if the number of missing studies, as indicated by the classic fail-safe N, exceeds the tolerance level of $5n + 10$. Here, "n" represents the original number of studies extracted without duplication ($n = 472$). Accordingly, according to the classic fail-safe N method applied to our study, an additional 3554 new results would need to be included to render the overall effect size statistically insignificant. This suggests a considerable degree of robustness in the findings against potential publication bias. This number is greater than the tolerance level value of 2370 [i.e., $5(472) + 10$], indicating the publication bias to be negligible and acceptable to run this meta-analysis.

Test for sensitivity. In a meta-analysis, sensitivity tests are frequently utilized to assess the resilience of the overall conclusions. Sensitivity analysis aims to identify any effect size that might exert

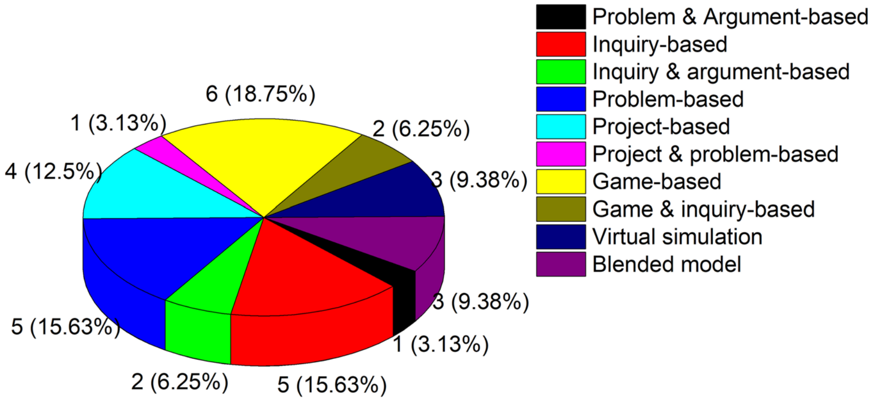

Fig. 4 Types of pedagogical models in high school biology. The pie chart reveals the pedagogical model types in high school biology by the number of studies included.

an undue impact on the central tendency (overall effect size) and variability of the data. This often arises from highly influential effect sizes (studies) located at the extremes of the distribution. Hence, the effect sizes of the studies were assessed using the "one study removed procedure" within the CMA software (Borenstein et al., 2009). The findings indicated that, with each study removed, the highest mean in the random model was $d = 0.740$ ($n = 32$, SE $= 0.089$), while the lowest mean was $d = 0.689$ ($n = 32$, SE $= 0.086$). Remarkably, both of these new average effect sizes fell within the confidence interval of the complete dataset, which was [$n = 32$, $d = 0.718$, 95% CI {0.547; 0.890}, $p < 0.001$]. This suggests that no anomalies were observed to exert a substantial influence on the calculated average ESs.

Results

All the finalized studies included in this study were reviewed and confirmed for pedagogies employed in a mixed-ability high school biology classroom. Using the meta-analysis, the effect sizes based on Cohen's d (standardized difference in means) were estimated, to assess the effectiveness of each type of pedagogical approach, considering the learning gains.

What are the various non-traditional pedagogical models/ approaches employed in high school biology education? To investigate the pedagogical approach employed in high school biology education, the methodology section of the studies included was thoroughly screened. The studies ($n = 32$) included in the paper were categorized based on the type of pedagogical model/approach employed. The following categories were identified initially: (a) game-based ($n = 6$), (b) inquiry-based ($n = 5$), (c) problem-based ($n = 5$), (d) project-based ($n = 4$), (e) virtual laboratory & simulation-based ($n = 3$), (f) blended model ($n = 3$). Some combinations were also classified (g) game & inquiry-based ($n = 2$), (h) argumentation & inquiry-based ($n = 2$), (i) project & problem-based ($n = 1$), (j) problem & argument-based ($n = 1$). Figure 4 represents a pie chart showing the pedagogical approach by studies.

What is the overall impact of the non-traditional pedagogical models employed for mixed-ability high school biology classrooms (when compared to the traditional lecture model)? The measure of effectiveness (effect size) has been determined by Cohen's d value. Indeed, Cohen's d value represents the standardized difference between means. Cohen's d is commonly used in scenarios where the independent variable is binary, and the dependent variable is continuous. A positive effect size would favor the non-traditional intervention, while a negative effect size would favor the conventional teaching model. Additionally, Cohen's d values of ≤0.2, 0.2–0.5, 0.5–0.7, and ≥0.8 are classified as indicating low, medium, high, and very high effects, respectively. The larger the Cohen's d values, the greater the mean difference compared to the variability, indicating greater reliability of the study findings. A random-effect model has been employed (refer to the "Methods" section). The overall effect size of non-traditional pedagogical approaches for mixed-ability high school biology classrooms, when compared to the traditional model has shown a high impact with an effect size value of 0.718. Figure 5 depicts a forest plot that has been used to reveal the distribution of effect sizes [$n = 32$, $d = 0.718$, 95% CI {0.547; 0.890}, $p < 0.001$]. In other words, non-traditional pedagogical models for mixed-ability high school biology classrooms are more highly effective than traditional methods. In the forest plot, the squares on the right represent the effect size of individual studies, while the diamond at the bottom illustrates the overall effect. The lines extending from the squares and diamonds indicate the confidence intervals. The range of effect sizes varied from low ($d = 0.012$) to high ($d = 3.15$) effects. Importantly, all the studies demonstrated a positive effect size, indicating a beneficial impact of the interventions. While only 5 among the 32 studies revealed a low effect size i.e., $d < 0.2$ (Mulder et al., 2016; Klisch et al., 2012; Su et al., 2014; Sivia et al., 2019a; Ebrahim and Naji, 2021). The heterogeneity analysis has estimated a Q value of 297.831 and an I^2 index of 89.59%. The distribution of true effects was also determined using CMA prediction interval software. This analysis indicates that the true effect sizes in 95% of all comparable populations are expected to fall within the prediction interval range of −0.21 to 1.69. This provides a range within which the true effects are likely to lie across different populations.

What is the comparative effectiveness of each pedagogical approach in improving students' gains (cognitive, affective, behavior) in mixed-ability high school biology classrooms? To investigate the effectiveness of the various pedagogical models, a subgroup meta-analysis was performed. The studies were classified into six pedagogical categories; (1) project-based; (2) problem-based; (3) inquiry-based; (4) blended model; (5) game-based; (6) virtual simulation-based (Table 2). Studies employing a combination of pedagogical models/ approaches were classified, as per Table 2 for the sub-group meta-analysis. The findings from the table reveal that the problem-based, inquiry-based, game-based, and argumentation-based pedagogies sought to improve students' cognitive, affective, and behavioral gains. While the studies incorporating the project-based models reported only

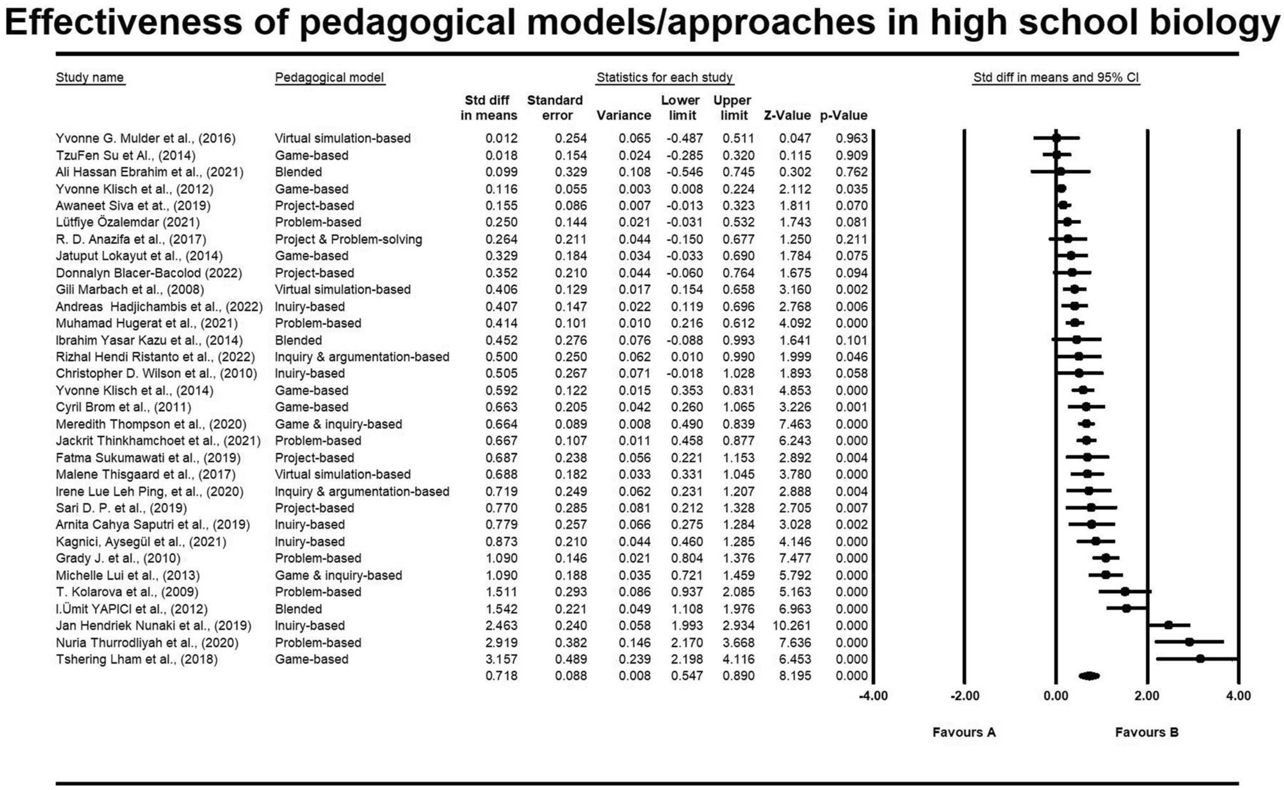

Fig. 5 Forest plot diagram depicting the distribution of effect sizes for the included studies. The squares depict effect sizes from individual studies, with the diamond representing the overall effect; lines extending from both indicate confidence intervals.

cognitive and behavioral gains. Similarly, the virtual simulation-based educational models contributed only toward cognitive and affective gains. These findings postulate the need for more research on project-based and virtual simulation-based educational models for high school biology, investigating the students' affective and behavioral gains respectively.

Figures 6–12 show the forest plot of the subgroup meta-analysis (comparative analysis). Wherein, the average effect size of project-based learning pedagogies has shown moderate effects on students' cognitive, and, behavioral gains ($n = 5$, $d = 0.374$, 95% CI [0.137; 0.612], $p = 0.002$, refer Fig. 6). The effect sizes ranged from non-significant small effect size ($d = 0.155$; Sivia et al., 2019a) to significant large effect size ($d = 0.77$; Sari et al., 2019).

The average effect size of problem-based learning pedagogies has also revealed a highly significant impact on students' cognitive, affective, and, behavioral gains ($n = 7$, $d = 0.913$, 95% *CI* [0.511; 1.315], $p < 0.001$, refer to Fig. 7). The effect sizes spanned from non-significant small effect size ($d = 0.250$; (Özalemdar, 2021)) to significant large effect size ($d = 2.919$; Thurrodliyah et al., 2020). As the I^2 value was reported to be greater than 75% (I^2-value = 91.3%); thus, it indicated that the large proportion of variability appears to be of true variance.

The random average effect size of the inquiry-based model has revealed a highly significant impact on students' cognitive, affective, and, behavioral gains ($n = 9$, $d = 0.882$, 95% CI [0.536; 1.227], $p < 0.001$, refer Fig. 8). The effect sizes varied from moderate effect size ($d = 0.500$; Ristanto et al., 2022) to significant large-sized effect ($d = 2.463$; Nunaki et al., 2019).

The overall effect size of blended learning approaches has shown high effectiveness on students' cognitive, affective, and, behavioral gains ($n = 3$, $d = 0.720$, 95% CI [−0.178; 1.619], $p = 0.116$, refer Fig. 9). The effect sizes varied from non-significant small effect size ($d = 0.099$; (Ebrahim and Naji, 2021)) to significant large effect size ($d = 1.542$; (Yapici and Akbayin, 2012). Finally, a comparative analysis have been performed and project-based learning is found to be most effective ($d = 0.913$) in augmenting learning gains in mixed-ability high school biology classrooms (Fig. 13).

The average effect size of game-based strategies has revealed a high impact on improving students' cognitive, affective, and, behavioral gains ($n = 8$, $d = 0.662$, 95% CI [0.335; 0.988], $p < 0.001$, refer to Fig. 10). The effect sizes ranged from low effect size ($d = 0.018$; Su et al., 2014) to significantly large effect size ($d = 3.157$; (Lham and Sriwattanarothai, 2018)). Among the 8 shortlisted studies employing a game-based model, 2 studies showed a non-significant p value (Su et al., 2014; Lokayut and Srisawasdi, 2014). However, the results are reliable as the average $p < 0.001$ with an average effect size of 0.662.

The random average effect size of virtual simulation-based models in high school biology education has shown moderate impacts in terms of students' cognitive, and, affective gains ($n = 3$, $d = 0.407$, 95% CI [0.089; 0.724], $p = 0.012$, refer Fig. 11). The effect sizes varied from non-significant small effect size ($d = 0.012$; Mulder et al., 2016) to significant large effect size ($d = 0.688$; Thisgaard and Makransky, 2017).

The average effect size of argumentation-based pedagogies in high school biology education has revealed high impacts in terms of students' gains ($n = 3$, $d = 0.815$, 95% CI [0.444; 1.185], $p = 0.010$, refer to Fig. 12). The effect sizes ranged from moderate effect size ($d = 0.500$; Ristanto et al., 2022) to significant large effect size ($d = 1.090$; (Venville and Dawson, 2010)).

Perspective

With the growing demand for personalization and improving students' learning gains, diverse non-traditional pedagogical

Table 2 Types of pedagogical models employed in high school biology by studies.

Study	Pedagogical Model	Sub-model/approach	Students' learning gains	Outcome measures	Effect size (Average)
Project-based					
Sivia et al. (2019a)	Project-based	-	B	Pre-post-test on on-task behavior (Chi-square analysis)	0.374 (moderate)
Sukmawati et al. (2019)	Project-based	Collaborative learning	C & B	Post-tests on social skills, knowledge (MANOVA)	
Sari et al. (2019)	Project-based	-	**C & B	Post-test on research skills test (Mann Whitney test)	
Blacer-Bacolod (2022)	Project-based	Video making	A, B, & C	Pre-post-test on engagement, knowledge, self-efficiency, and technical skills (*t* test analysis)	
Anazifa and Djukri (2017)	Project & problem-based	-	**C & B	Pre-post-test of creativity & and critical thinking skills (*t* test analysis)	
Problem-based					
Kolarova et al. (2014)	Problem-solving model	Reflexive model	A, B, & C	Pre-post-test on Levels of reflection (*t* test analysis)	0.913 (very high)
Özalemdar (2021)	Problem-solving model	Active learning method	**C & B	Pre-post-test on environmental thinking and behavior (*t* test analysis)	
Hugerat et al. (2021)	Problem-solving model	Problem-based jigsaw discussion	*C & B	Pre-post-test on knowledge, motivation, and classroom climate (*t* test analysis).	
Venville and Dawson (2010)	Problem & argument-based	Classroom-based argumentation	**C & B	Pre-post-test on knowledge, argumentation skills, (*t* test analysis)	
Thinkhamchoet et al. (2021)	Project-based	Team activity-based learning	**C & B	Pre-post-test on Environmental knowledge, ethics (variance analysis)	
Thurrodliyah et al. (2020)	Problem-solving model	Brain-based model	A, B, & C	Pre-post-test of cognitive, psychomotor, and, critical thinking skills (*t* test analysis)	
Anazifa and Djukri (2017)	Project & problem-based	Project-based and problem-based	**C & B	Pre-post-test on creativity, and, critical thinking skills (*t* test analysis)	
Inquiry-based					
Lui and Slotta (2014)	Inquiry & game-based	Immersive inquiry-based simulation	**A, B & C	Students' observations, pre-post-test on activity and knowledge (*t* test analysis)	0.882 (very high)
Thompson et al. (2020)	Inquiry & game-based	Inquiry-based 3-D virtual reality	**C	Pre-post-test on knowledge (*t* test analysis)	
Saputri et al. (2019)	Inquiry-based	Higher-order thinking model	C	Pre-post-test on creative thinking skills (*t* test analysis)	
Hadjichambis et al. (2022)	Inquiry-based	Environmental model	**A, B & C	Pre-post-test on Environmental Knowledge, conception, attitude, skills, values (Effect size calculated)	
Nunaki et al. (2019)	Inquiry-based	Materials-oriented inquiry-based	*C	Pre-post-test of meta-cognitive skills (*t* test analysis)	
Kagnici and Sadi (2021)	Inquiry-based	5 E Model	**A, B & C	Pre-post-test on content knowledge, MANOVA)	
Wilson et al. (2010)	Inquiry-based	5 E Model	**A, B & C	Pre-post test on knowledge, reasoning, and, argumentation skills (effect size analysis)	
Ristanto et al. (2022)	Inquiry & argument based	Guided discovery-argument model	**C	Pre-post-test on critical thinking skills (*t* test analysis)	
Ping, et al. (2020)	Inquiry & argument based	Argumentation-based	**C	Pre-post-test on knowledge argumentation skills (MANOVA)	
Blended model					
Ebrahim and Naji (2021)	Blended model	Flipped learning	C	Pre-post-test on social intelligence, and, biology attainment (*t* test analysis)	0.720 (high)
Yapici and Akbayin (2012)	Blended model	-	**C & A	Pre-post-test on Knowledge, attitude, (*t* test analysis)	
Kazu and Demirkol (2014)	Blended model	-	C	Pre-post-test on Knowledge, attitude, (*t* test analysis)	
Game-based					
Su et al. (2014)	Game-based	Educational card game	C & A	Pre-post-test on student perception of instruction, learning efficiency	0.662 (high)
Klisch et al. (2012)	Game-based	Online game-based	**A, B, & C	Pre-post-test on Knowledge, attitude, and, satisfaction (regression analysis)	
Lham and Sriwattanarothai (2018)	Game-based	Boardgame	C	Pre-post-test on knowledge attainment (*t* test analysis)	
Brom et al. (2011)	Game-based	Online micro game	*A	Pre-post-test on knowledge attainment (effect-size analysis)	

Table 2 (continued)

Study	Pedagogical Model	Sub-model/approach	Students' learning gains	Outcome measures	Effect size (Average)
Lokayut and Srisawasdi (2014)	Game-based	Computer game-based	**C & A	Pre-post-test on knowledge, perception, and self-confidence (*t* test analysis)	
Klisch et al. (2013)	Game-based	Online game-based	**C & A	Pre-post-test on Knowledge, attitude, and, satisfaction (regression analysis)	
Lui and Slotta (2014)	Game & inquiry-based	Immersive inquiry-based simulation	**A, B, C	Students' observations, pre-post-test on knowledge (*t* test analysis)	
Thompson et al. (2020)	Game & inquiry-based	Inquiry-based 3-D virtual reality	**C	Pre-post-test on knowledge (*t* test analysis)	
Virtual simulation-based					
Thisgaard and Makransky (2017)	Virtual simulation	Virtual laboratory simulation	A, B, & C	Multiple regression model for interest, self-efficiency, motivation, outcome expectations in biology	0.407 (Moderate)
Mulder et al. (2016)	Virtual simulation	Online scientific modeling	**C	Pre-post-test concept knowledge, reasoning knowledge (MANOVA)	
Marbach-Ad et al. (2008)	Virtual simulation	Computer animation, and, illustration	**C	Pre-post-test on knowledge (MANOVA)	
Argument-based					
Venville and Dawson (2010)	Problem & argument-based	Classroom-based argumentation	**C & B	Pre-post-test on knowledge, argumentation skills, (*t* test analysis)	0.815 (very high)
Ping, et al. (2020)	Inquiry & argument-based	-	**C & B	Pre-post-test on knowledge argumentation skills (MANOVA)	
Ristanto et al. (2022)	Inquiry & argument-based	Guided discovery argument model	**C	Pre-post-test on critical thinking skills (*t* test analysis)	

A affective learning gains, *B* behavioral learning gains, *C* cognitive learning gains, *MANOVA* multivariate analysis of variance.
*Significant at 0.05 level. **Significant at 0.05 level and Cohen's $d \geq 0.4$.

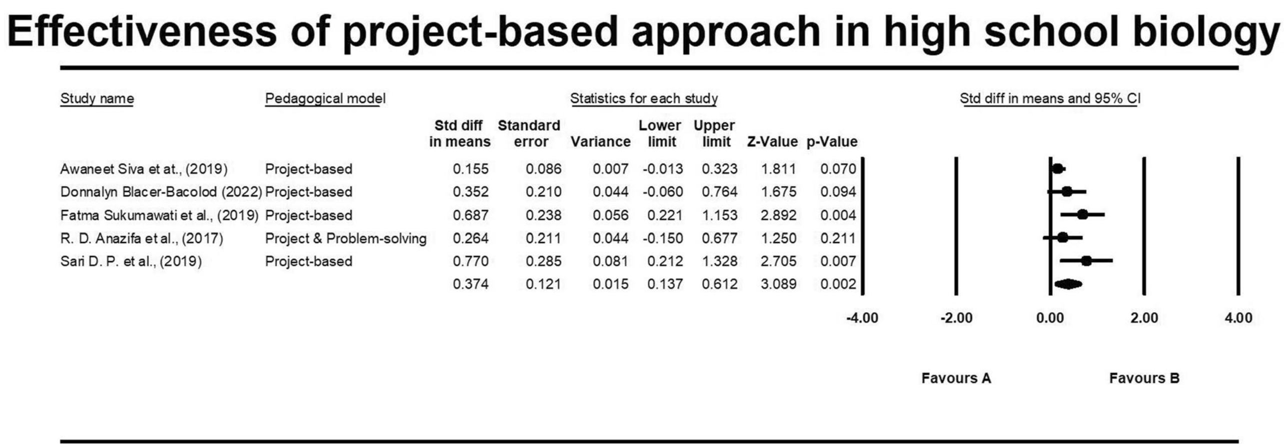

Fig. 6 Forest plot depicting the distribution of effect size values for studies employing the project-based pedagogical approach. The squares depict effect sizes from individual studies, with the diamond representing the overall effect; lines extending from both indicate confidence intervals.

practices are developed, tested, and employed by researchers and academicians. We believe that for a satisfactory, and, suitable approach, a dynamic model/ approach must be utilized for varied student populations/communities. Therefore, our first objective of the study was to investigate the diverse pedagogical models/ approaches employed in high school biology education. Some of the non-traditional models employed for biology education have been reported to be virtual laboratory simulation-based, inquiry-based, argumentation-based, problem-based, project-based, game-based, blended models, etc. (Klisch et al., 2013; Nunaki et al., 2019; Ping et al., 2020; Sivia et al., 2019a; Thisgaard and Makransky, 2017; Thurrodliyah et al., 2020; Yapici and Akbayin, 2012). Furthermore, a meta-analysis review of the existing literature has been performed to investigate the impact of the non-traditional pedagogical approaches on mixed-ability high school biology classrooms. The effectiveness was revealed by the standardized difference in means (Cohen's *d*). The overall impact of the non-traditional intervention was found to be significantly effective [$n = 32$, $d = 0.809$, 95% CI {0.604; 1.015}, $p < 0.001$] when compared to traditional lectures. A subgroup meta-analysis of various non-traditional pedagogical approaches has also been performed to reveal the pedagogical comparative assessment. Figure 5 reports the overall random average effect sizes of the various non-traditional pedagogies. Furthermore, results from moderator analysis reveal that there is no influence of "biology topics" on the effect size of the studies (due to in-significant Q_B statistics: between-class variance component statistics).

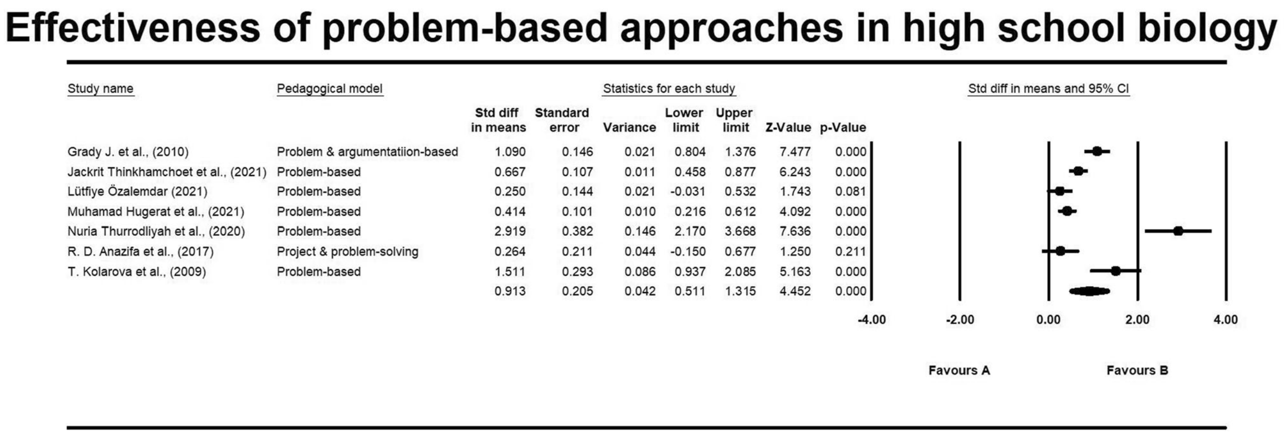

Fig. 7 Forest plot depicting the distribution of effect size values for studies employing the problem-based pedagogical approach. The squares depict effect sizes from individual studies, with the diamond representing the overall effect; lines extending from both indicate confidence intervals.

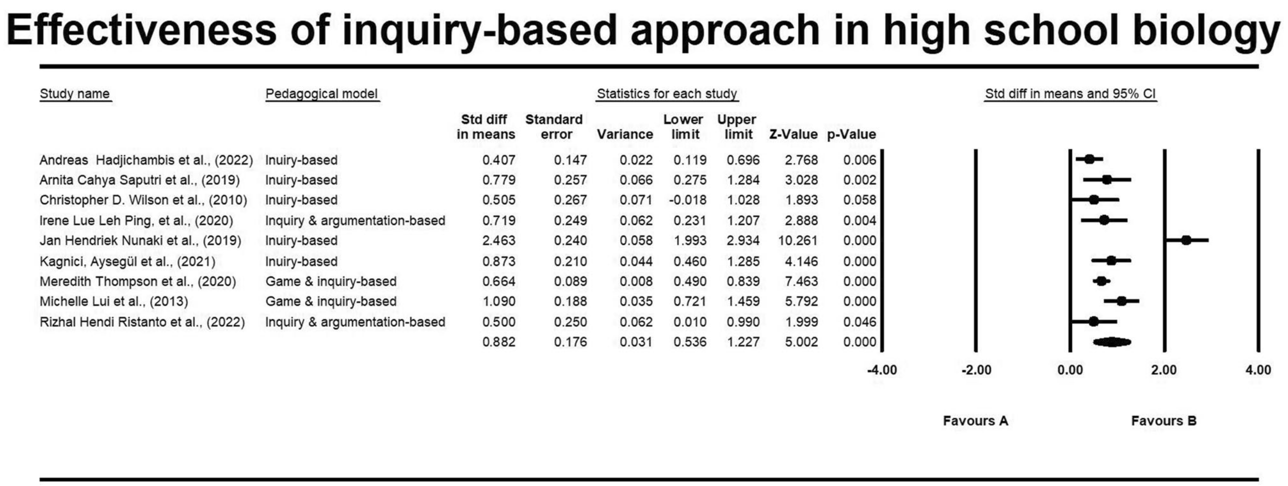

Fig. 8 Forest plot depicting the distribution of effect size values for studies employing the inquiry-based pedagogical approach. The squares depict effect sizes from individual studies, with the diamond representing the overall effect; lines extending from both indicate confidence intervals.

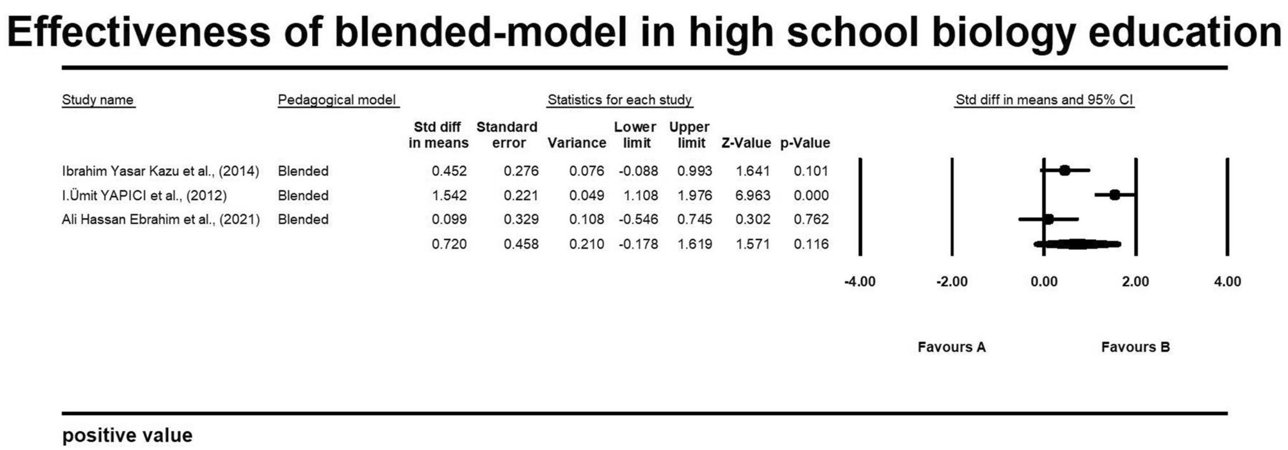

Fig. 9 Forest plot depicting the distribution of effect size values for studies employing the blended pedagogical approach. The squares depict effect sizes from individual studies, with the diamond representing the overall effect; lines extending from both indicate confidence intervals.

The problem-based pedagogical model was found to be highly effective in augmenting students' cognitive, and, behavioral gains [$n = 7$, $d = 0.913$, 95% CI {0.511; 1.315}, $p = 0.00$]. These results are in line with the study findings by Xu et al. (2021), who also reported high effectiveness of the problem-based model with Cohen's d value ≈ 0.9 for biology education (Xu et al., 2021). It is important to note that no studies in this category reported significant affective gains. This low effectiveness could be due to the

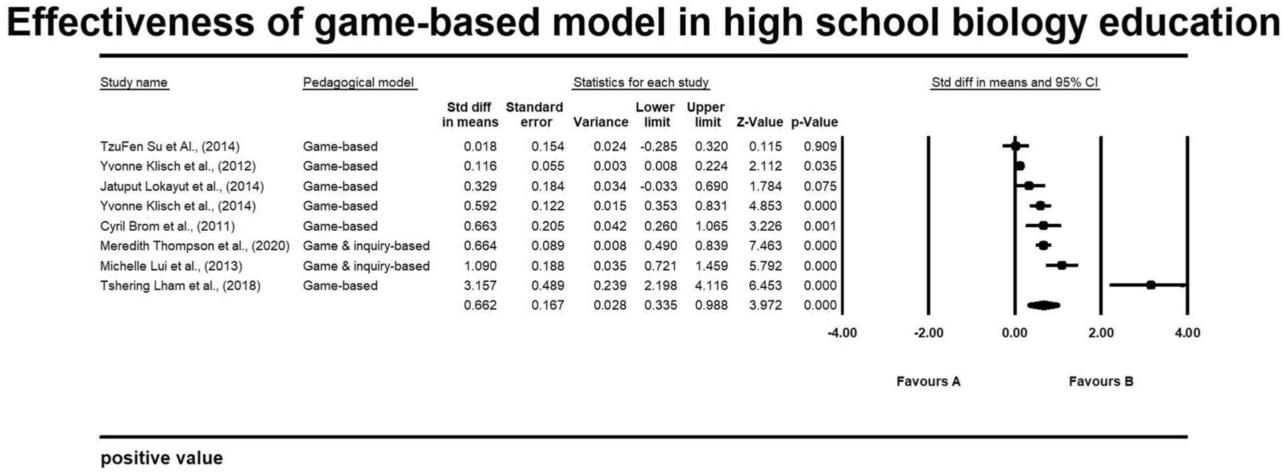

Fig. 10 Forest plot depicting the distribution of effect size values for studies employing the game-based pedagogical approach. The squares depict effect sizes from individual studies, with the diamond representing the overall effect; lines extending from both indicate confidence intervals.

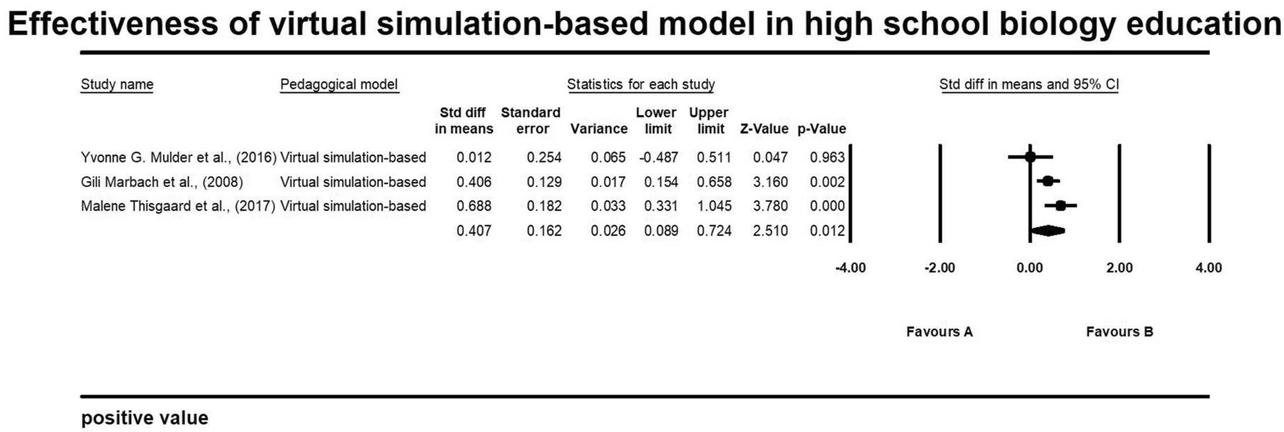

Fig. 11 Forest plot depicting the distribution of effect size values for studies employing the virtual simulation-based pedagogical approach. The squares depict effect sizes from individual studies, with the diamond representing the overall effect; lines extending from both indicate confidence intervals.

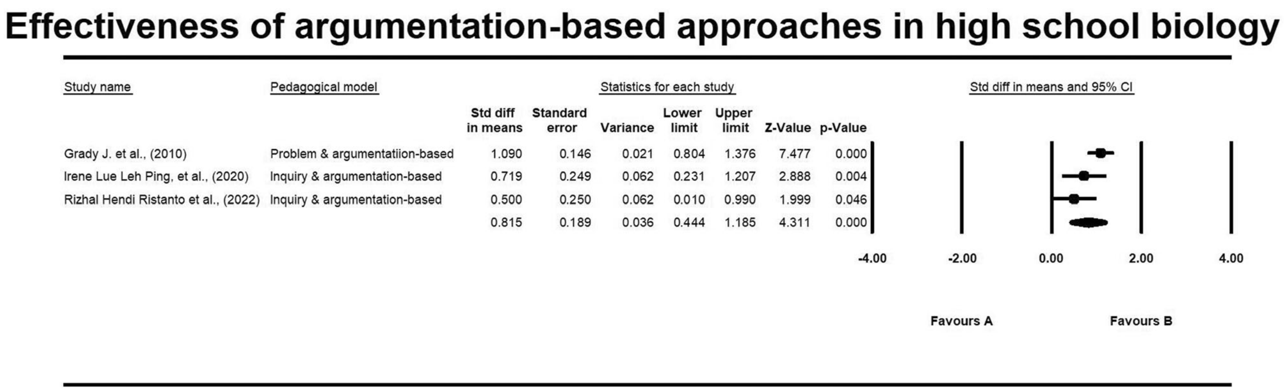

Fig. 12 Forest plot depicting the distribution of effect size values for studies employing the argumentation-based pedagogical approach. The squares depict effect sizes from individual studies, with the diamond representing the overall effect; lines extending from both indicate confidence intervals.

limitations of the problem-based model. Which includes devoting greater time and effort to problem-based learning might cause low performance in standardized assessments, as the students might have been involved actively and acquired sufficient knowledge but might not have the depth of knowledge required to score well. Further analysis showcased that students' affective outcomes are often ignored in literature while reporting/assessing the problem-based models in biology education.

On investigating the inquiry-based learning pedagogies, it has been reported to be highly impactful on students' cognitive, affective, and, behavioral learning gains ($n = 9$, $d = 0.882$, 95% CI [0.536; 1.227], $p < 0.001$, refer Fig. 8). All the studies employing

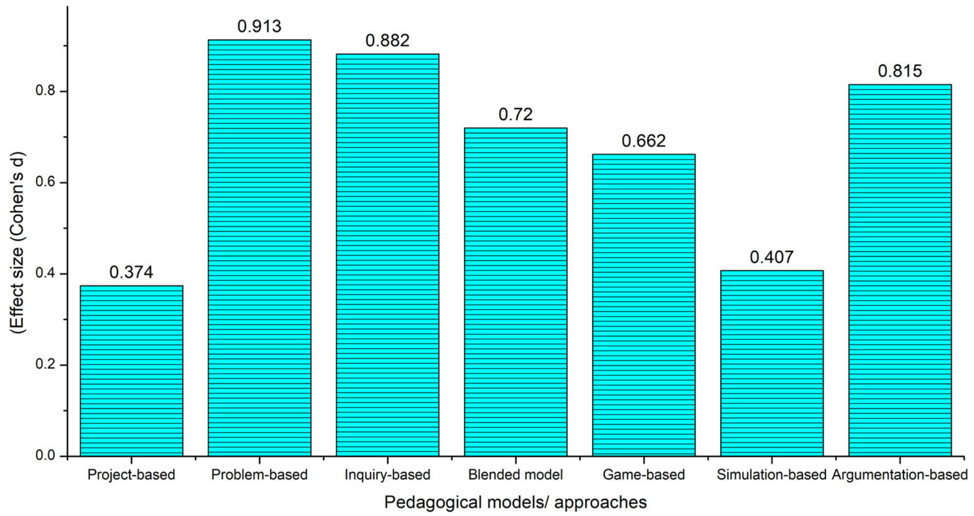

Fig. 13 Comparative analysis of effect sizes. The bar graph represents the effect sizes (Cohen's *d*) of various pedagogical approaches in high school biology education.

an inquiry-based model were statistically significant and impactful. These findings are partially in agreement with the study by Funa and Prudente (2021), which reported the impact of inquiry-based pedagogies for biology education to be Cohen's d value $= 1.26$ (Funa and Prudente, 2021). The biological science inquiry model is one of the progressing approaches that aid students in processing information using various techniques used by biologists (Furtak et al., 2012). In this area, the students/researchers try to identify different problems and use scientific methodology to solve the problems (Funa and Prudente, 2021). The reported I^2 value was greater than 75% (I^2-value = 87.2%); indicating a larger proportion of variability to be true variance.

Similarly, the average effect size of argumentation-based pedagogies in high school biology education has been reported to be highly impactful in terms of students' cognitive, affective, and, behavioral learning gains ($n = 3$, $d = 0.815$, 95% CI [0.444; 1.185], $p = 0.010$, refer Fig. 12). Argumentation in biological science teaching is demonstrated as a process of scientific investigation that involves the justification of claims based on evidence. Thus, by acquiring argumentation skills the student could develop science process skills (Ping et al., 2020; Ristanto et al., 2022). A possible strategy to acquire argumentation skills in the science classroom is by performing laboratory investigations. Wherein, the students will be involved in investigating, tabulating, and analyzing data systematically, thus producing evidence to defend the claims (Ping et al., 2020). Further investigation revealed that students' affective outcomes are often ignored in literature while evaluating the argumentation-based models in biology education (Ping et al., 2020, Ristanto et al., 2022; Venville and Dawson, 2010).

Followingly, the blended and game-based models have revealed a high effect size of 0.720 and 0.662 respectively. Even though the average effect size of blended learning pedagogies has revealed high effectiveness on students learning ($n = 3$, $d = 0.720$, 95% CI [−0.178; 1.619], $p = 0.116$, refer Fig. 9). There were only 3 studies employing a blended model and the overall p-value is not significant. Therefore, more studies might be included to conclude and justify its impact on students. Moreover, among these three studies, behavioral gains were neglected and not evaluated. Followingly, in the context of game-based pedagogies, findings have illustrated a high impact in terms of students learning ($n = 8$, $d = 0.662$, 95% *CI* [0.335; 0.988], $p < 0.001$, refer to Fig. 10).

Finally, on investigating the effect of project-based and virtual simulation-based models, their effect sizes are found to be moderately impactful (effect sizes 0.374 and 0.0.407 respectively). Wherein, the average effect size of project-based learning pedagogies has shown moderate effects on students' cognitive, and, behavioral gains ($n = 5$, $d = 0.374$, 95% CI [0.137; 0.612], $p = 0.002$, refer Fig. 6). These results do not align with the previous meta-analytical research where project-based learning in biology education has reported greater effect sizes [i.e., $d = 1.36$ (Balemen and Keskin, 2018); $d = 0.95$ (Ayaz and Söylemez, 2015)]. Despite the project-based model showing its effectiveness, the studies have only reported/assessed cognitive and behavioral gains, giving meager importance to affective gains. Followingly, in the context of virtual simulation-based pedagogies, findings revealed moderate impacts in terms of students' cognitive learning gains ($n = 3$, $d = 0.407$, 95% CI [0.089; 0.724], $p = 0.012$, refer to Fig. 11). These results are inconsistent with prior meta-analyses reporting a greater effect size of virtual simulation-based models for biology education [$d = 0.668$; Bayraktar, 2001]. It is also noteworthy that only cognitive gains were reported/assessed in the shortlisted studies, neglecting the affective and behavioral gains.

Furthermore, Table 3 briefs the comparative aspect along with the heterogeneity analysis of the sub-groups. Cohen's d value determining the effect sizes of the sub-group pedagogical models has revealed very high effects for argumentation-based ($d = 0.815$), problem-based ($d = 0.913$), and inquiry-based ($d = 0.882$). And high effects for blended ($d = 0.720$) and game-based models ($d = 0.662$). And moderate impacts for virtual simulation ($d = 0.407$) and project-based ($d = 0.374$)] on students' gains. However, among all the sub-group analyses, only the p value for the blended model is non-significant (p value = 0.116). The I^2 value is highly significant for problem-based, inquiry-based, game-based, and blended models (I^2-value = 91.3%, 87.2%, 92.21%, and 88.21% respectively). The higher I^2 value has revealed that a large proportion of variability is due to true variance. Likewise, the argumentation-based, virtual simulation-based, and project-based have also shown significant I^2-value = 58.23%, 57.83%, and 50.71% respectively.

In conclusion, it is important to interpret the findings of this study with an awareness of certain limitations. These encompass the absence of a discussion on the feasibility of non-traditional

Table 3 Comparative analysis of the pedagogical subgroups: heterogeneity statistics.

Study	Average	95% Confidence interval			Heterogeneity		
	Cohen's *d*	Lower limit	Upper limit	*p* value	*Q*-value	Df	I^2 value
Project-based	0.374	0.137	0.612	0.002	8.115	4	50.71%
Problem-based	0.913	0.511	1.315	0.000	69.52	6	91.37%
Inquiry-based	0.882	0.536	1.227	0.000	62.54	8	87.20%
Blended	0.720	−0.178	1.619	0.116	16.96	2	88.21%
Game-based	0.662	0.335	0.988	0.000	89.90	7	92.21%
Virtual simulation	0.407	0.089	0.724	0.012	4.74	2	57.83%
Argument-based	0.815	0.444	1.185	0.000	4.78	2	58.23%

Note: d = Cohen's d (standardized difference in mean); Q-value = Cochran's Q; Df = degree of freedom.

pedagogical approaches, as these practices can often be constrained by factors such as implementation costs, and the time and effort involved. Additionally, it's worth noting that our study exclusively incorporated peer-reviewed articles, with data from theses, books, and proceedings excluded from consideration. Including such gray literature could have given a holistic view of the available evidence. Another limitation is the inability of study findings to be expressed in terms of different types of learning gains (i.e., creativity, scientific thinking skills, etc.) under a specific pedagogical category due to the lack of studies. Rather authors could infer if cognitive, affective, and behavioral gains are targeted in each pedagogical model.

Conclusion

The main contribution of this study is the analysis of the impact of non-traditional pedagogical models/approaches employed for mixed-ability high school biology classrooms. Considering the diversity in students' needs, and with the growing demand for personalization and improving students' learning gains (cognitive, affective, and behavioral), diverse teaching practices are developed, tested, and employed by researchers and academicians. Some of the non-traditional models employed for biology education are virtual laboratory simulation-based, inquiry-based, argumentation-based, problem-based, project-based, game-based, blended model, etc. This meta-analysis review has reported a high impact of the non-traditional pedagogical approach in comparison to the traditional lecture method in mixed-ability high school biology classrooms [$n = 32$, $d = 0.809$]. In addition, a subgroup meta-analysis of various non-traditional pedagogical approaches revealing the comparative aspect reported a very high impact of the problem-based model for augmenting students' learning gains [$n = 7$, $d = 0.913$]. Further investigation showcases that it is cognitive gains that are often explored, followed by behavioral, and least by affective gains in mixed-ability high school biology classrooms. Therefore, this study proposes the necessity for future studies evaluating affective gains during project-based, problem-based, and argumentation-based models and behavioral gains during blended models. Thus, these findings propose the necessity for further studies investigating affective and behavioral gains during project-based and virtual simulation-based models respectively.

Acknowledging the limitations discussed in the previous section, our study highlights several promising avenues for future research. As a result, it offers valuable insights for researchers engaged in the study of non-traditional pedagogies. Subsequent studies could delve into the necessity for longitudinal pedagogical interventions, recognizing that certain student variables may require time to manifest and subsequently be assessed. This can contribute to a deeper understanding of the long-term impacts of such teaching methods. Studies on how low-competency students perform differently than high-competency students during/after the pedagogical intervention could be researched. The future scope of the study also includes various moderators' analyses, such as grades of the participants, duration of the intervention, gender of participants, knowledge type targeted, etc. The moderator analysis will aid in understanding the effect of pedagogy with respect to specific characteristics of the intervention. Thus, finally, we believe that this research would pave the way for academicians to design, customize, and implement novel pedagogies for a dynamic education system in high school biology classrooms (considering the learning requirements of the classrooms).

Data availability

All data generated or analyzed during this study are included in this published article. The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request from Havard Dataverse at https://doi.org/10.7910/DVN/DBWZQH.

Received: 15 August 2022; Accepted: 6 November 2023;
Published online: 26 January 2024

References

Anazifa RD, Djukri D (2017) Project- based learning and problem- based learning: Are they effective to improve student's thinking skills? J Pendidikan IPA Indonesia 6(2):346–355. https://doi.org/10.15294/jpii.v6i2.11100

Arnold I (2011) John Hattie: Visible learning: a synthesis of over 800 meta-analyses relating to achievement. Int Rev Educ. https://doi.org/10.1007/s11159-011-9198-8

Ayaz MF, Söylemez M (2015) The effect of the project-based learning approach on the academic achievements of the students in science classes in Turkey: a meta-analysis study. Egitim Bilim. https://doi.org/10.15390/EB.2015.4000

Bayraktar S (2001) A meta-analysis of the effectiveness of computer-assisted instruction in science education. J Res Technol Educ 34(2):173–188. https://doi.org/10.1080/15391523.2001.10782344

Balemen N, Keskin MÖ (2018) The effectiveness of Project-Based Learning on science education: A meta analysis search. Int Online J Educ Teach 5(4):849–865

Blacer-Bacolod D (2022) Student-generated videos using green screen technology in a biology class. Int J Inf Educ Technol 12(4):339–345. https://doi.org/10.18178/IJIET.2022.12.4.1624

Bloom BS, Englehart MD, Furst EJ, Hill WH, Krathwohl DR (1956) Taxonomy of educational objectives. Handbook I: Cognitive domain. David McKay Company, New York

Borenstein M, Hedges LV, Higgins JPT, Rothstein HR (2009) Introduction to meta-analysis. Wiley, p 1–421. https://doi.org/10.1002/9780470743386

Brom C, Preuss M, Klement D (2011) Are educational computer micro-games engaging and effective for knowledge acquisition at high-schools? A quasi-experimental study. Comput Educ 57(3):1971–1988. https://doi.org/10.1016/J.COMPEDU.2011.04.007

Cai Z, Mao P, Wang D, He J, Chen X, Fan X (2022) Effects of scaffolding in digital game-based learning on student's achievement: a three-level meta-analysis. Educ Psychol Rev 34(2):537–574

Chaplin S (2007) A model of student success: coaching students to develop critical thinking skills in introductory biology courses. Int J Scholarship Teaching Learn 1(2):n2. https://doi.org/10.20429/ijsotl.2007.010210

Cimer A (2012) What makes biology learning difficult and effective. Students' views. Educ Res Rev 7(3):61–71
Cochran WG (1954) Some methods for strengthening the common χ2 tests. Biometrics 10(4):417. https://doi.org/10.2307/3001616
Cohen J (2013) Statistical power analysis for the behavioral sciences. Routledge
Cronbach LJ, Furby L (1970) How we should measure "change": or should we? Psychol Bull 74(1):68–80. https://doi.org/10.1037/h0029382
DerSimonian R, Laird N (2015) Meta-analysis in clinical trials revisited. Contemp Clin Trials 45:139–145. https://doi.org/10.1016/j.cct.2015.09.002
Duval S, Tweedie R (2000) Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics 56(2):455–463. https://doi.org/10.1111/J.0006-341X.2000.00455.X
Ebrahim AH, Naji SAB (2021) The influence of flipped learning methods on high school learners' biology attainment and social intelligence in Kuwait. Eurasia J Math Sci Technol Educ 17(8):em1987. https://doi.org/10.29333/EJMSTE/10997
Funa AA, Prudente MS (2021) Effectiveness of problem-based learning on secondary students' achievement in science: a meta-analysis. Int J Instruct 14(4):69–84
Furtak EM, Seidel T, Iverson H, Briggs DC (2012) Experimental and quasi-experimental studies of inquiry-based science teaching: a meta-analysis. Rev Educ Res 82(3):300–329. https://doi.org/10.3102/0034654312457206
Goodwin B, Dawkins R (1995) What is an organism? A discussion. Perspectives in ethology 11:47–60
Hadjichambis AC, Paraskeva-Hadjichambi D, Georgiou Y(2022) Evaluating a novel learning intervention grounded in the education for environmental citizenship pedagogical approach: a case study from Cyprus Sustainability 14(3):1398. https://doi.org/10.3390/SU14031398
Hsu YC, Hung JL, Ching YH (2013) Trends of educational technology research: more than a decade of international research in six SSCI-indexed refereed journals. Educ Technol Res Dev 61:685–705. https://doi.org/10.1007/s11423-013-9290-9
Hugerat M, Kortam N, Kassom F, Algamal S, Asli S (2021) Improving the motivation and the classroom climate of secondary school biology students using problem-based – jigsaw discussion (PBL-JD) learning. Eurasia J Math Sci Technol Educ 17(12):1–12. https://doi.org/10.29333/EJMSTE/11304. em2036
Kagnici A, Sadi Ö (2021) Students' conceptions of learning biology and achievement after STEM activity enriched instruction. IE Inquiry Educ 13(1):7
Kazu IY, Demirkol M (2014) Effect of blended learning environment model on high school students' academic achievement. Turk Online J Educ Technol 13(1):78–87
Klisch Y, Bowling K, Miller L, Ramos M (2013) The impact of science education games on prescription drug abuse attitudes among teens: a case study. J Drug Educ 43(3):255–275. https://doi.org/10.2190/DE.43.3.d
Klisch Y, Miller LM, Beier ME, Wang S (2012) Teaching the biological consequences of alcohol abuse through an online game: impacts among secondary students. CBE Life Sci Educ 11(1):94–102. https://doi.org/10.1187/cbe.11-04-0040
Kolarova T, Hadjiali I, Vasilev V (2014) Reflective approach to studying of genetics in 9th-10th grade. Biotechnol Biotechnol Equip 23:53–57. https://doi.org/10.1080/13102818.2009.10818364
Korkor Sam C, Acheaw Owusu K, Anthony-Krueger C (2018) Effectiveness of 3E, 5E and conventional approaches of teaching on students' achievement in high school biology. Am J Educ Res 6(1):76–82. https://doi.org/10.12691/education-6-1-12
Lham T, Sriwattanarothai N (2018) A board game to enhance understanding of cell cycle for grade ten Bhutanese students. Rabsel 19(2):1–17
Li Q, Ma X (2010) A meta-analysis of the effects of computer technology on school students' mathematics learning. Educ Psychol Rev 22:215–243. https://doi.org/10.1007/s10648-010-9125-8
Lokayut J, Srisawasdi N (2014) Designing educational computer game for human circulatory system: a pilot study. Workshop Proceedings of the 22nd International Conference on Computers in Education, ICCE. Asia-Pacific Society for Computers in Education, p 571–578
Lui M, Slotta JD (2014) Immersive simulations for smart classrooms: exploring evolutionary concepts in secondary science. Technol Pedagogy Educ 23(1):57–80. https://doi.org/10.1080/1475939X.2013.838452
Marbach-Ad G, Rotbain Y, Stavy R (2008) Using computer animation and illustration activities to improve high school students' achievement in molecular genetics. J Res Sci Teach 45(3):273–292. https://doi.org/10.1002/tea.20222
Mulder YG, Bollen L, de Jong T, Lazonder AW (2016) Scaffolding learning by modelling: the effects of partially worked-out models. J Res Sci Teach 53(3):502–523. https://doi.org/10.1002/tea.21260
Nunaki JH, Damopolii I, Kandowangko NY, Nusantari E (2019) The effectiveness of inquiry-based learning to train the students' metacognitive skills based on gender differences. Int J Instr 12(2):505–516. https://doi.org/10.29333/iji.2019.12232a
Ostrom TM (1969) The relationship between the affective, behavioral, and cognitive components of attitude. J Exp Soc Psychol. https://doi.org/10.1016/0022-1031(69)90003-1
Özalemdar L (2021) The effect on environmental attitude of the active learning method applied in teaching the biology topic "Current Environmental Issues and Human" for 10th grade students. J Turk Sci Educ 18(2):276–289. https://doi.org/10.36681/tused.2021.65
Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Moher D (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Syst Rev 10(1):1–11. https://doi.org/10.1136/BMJ.N71
Pascarella ET, Blaich C (2013) Lessons from the Wabash national study of liberal arts education. Change Mag High Learn 45(2):6–15. https://doi.org/10.1080/00091383.2013.764257
Ping ILL, Halim L, Osman K (2020) Explicit teaching of scientific argumentation as an approach in developing argumentation skills, science process skills and biology understanding. J Baltic Sci Educ 19(2):276–288. https://doi.org/10.33225/jbse/20.19.276
Reed DA, Cook DA, Beckman TJ, Levine RB, Kern DE, Wright SM (2007) Association between funding and quality of published medical education research. JAMA 298(9):1002–1009. https://doi.org/10.1001/jama.298.9.1002
Ristanto RH, Ahmad AS, Komala R (2022) Critical thinking skills of environmental changes: a biological instruction using guided discovery learning-argument mapping (GDL-AM). Particip Educ Res 9(1):173–191. https://doi.org/10.17275/PER.22.10.9.1
Roohr KC, Liu H, Liu OL (2017) Investigating student learning gains in college: a longitudinal study. Stud High Educ 42(12):2284–2300. https://doi.org/10.1080/03075079.2016.1143925
Rosenthal R (1979) The file drawer problem and tolerance for null results. Psychol Bull 86(3):638–641. https://doi.org/10.1037/0033-2909.86.3.638
Saputri AC, Sajidan RY, Prasetyanti NM (2019) Improving students' critical thinking skills in cell-metabolism learning using stimulating higher order thinking skills model. Int J Instr 12(1):327–342
Sari DP, Wulan AR, Solihat R (2019) Developing 21st-century student research skills through assessment matrix and edmodo in biology project. J Phys Conf Ser 1157(2):1–5. https://doi.org/10.1088/1742-6596/1157/2/022093
Sivia A, MacMath S, Novakowski C, Britton V (2019a) Examining student engagement during a project-based unit in secondary science. Can J Sci Math Technol Educ 19(3):254–269. https://doi.org/10.1007/S42330-019-00053-X/TABLES/8
Smith RP, Learman LA (2017) A plea for MERSQI: the medical education research study quality instrument. Obstet Gynecol 130(4):686–690. https://doi.org/10.1097/AOG.0000000000002091
Steenbergen-Hu S, Olszewski-Kubilius P, Calvert E (2020) The effectiveness of current interventions to reverse the underachievement of gifted students: findings of a meta-analysis and systematic review. Gifted Child Q 64(2):132–165. https://doi.org/10.1177/0016986220908601
Su TF, Cheng MT, Lin SH (2014) Investigating the effectiveness of an educational card game for learning how human immunology is regulated. CBE Life Sci Educ 13(3):504–515. https://doi.org/10.1187/CBE.13-10-0197/ASSET/IMAGES/LARGE/504FIG4.JPEG
Sukmawati F, Setyosari P, Sulton S, Purnomo P (2019) The effect of project-based collaborative learning and social skills on learning outcomes in biology learning. J Educ Gifted Young Sci 7(4):1325–1344. https://doi.org/10.17478/jegys.630693
Thinkhamchoet J, Wongchantra P, Bunnaen W (2021) The effects of environmental conservation school activities development using team work based learning (TWBL) for students of MuengRoi-Et municipality schools, Thailand. Ann RomSoc Cell Biol 25(5):5448–5464
Thisgaard M, Makransky G (2017) Virtual learning simulations in high school: effects on cognitive and non-cognitive outcomes and implications on the development of STEM academic and career choice. Front Psychol 8:805. https://doi.org/10.3389/FPSYG.2017.00805/BIBTEX
Thompson MM, Wang A, Uz Bilgin C, Anteneh M, Roy D, Tan P, … Klopfer E (2020) Influence of virtual reality on high school students' conceptions of cells. J Univers Comput Sci 26:926–946. Retrieved from https://dspace.mit.edu/handle/1721.1/130512
Thurrodliyah NI, Prihatin J, Novenda IL (2020) The development of brain-based learning model based on socio-scientific issues (Bbl-Ssi) for biology learning in senior high school. ScienceEdu 3(1):32–42. https://doi.org/10.19184/se.v3i1.17697
Yapici IU, Akbayin H (2012) The effect of blended learning model on high school students' biology achievement and on their attitudes towards the internet. Turk Online J Educ Technol 11(2):228–237
Venville GJ, Dawson VM (2010) The impact of a classroom intervention on grade 10 students' argumentation skills, informal reasoning, and conceptual understanding of science. J Res Sci Teach 47(8):952–977. https://doi.org/10.1002/TEA.20358
Wang JR, Huang BY, Tsay RF, Lee KP, Lin SW, Kao HL (2011) A meta-analysis of inquiry-based instruction on student learning outcomes in Taiwan. Asia Pac Educ Res 20(3):534–542
Warrens MJ (2015) Five ways to look at Cohen's kappa. J Psychol Psychother 5(4):1. https://doi.org/10.4172/2161-0487.1000197

Wei X, Saab N, Admiraal W (2021) Assessment of cognitive, behavioral, and affective learning outcomes in massive open online courses: a systematic literature review. Comput Educ 163:104097. https://doi.org/10.1016/j.compedu.2020.104097
Wilson CD, Taylor JA, Kowalski SM, Carlson J (2010) The relative effects and equity of inquiry-based and commonplace science teaching on students' knowledge, reasoning, and argumentation. J Res Sci Teach 47(3):276–301. https://doi.org/10.1002/tea.20329
Xu W, Ye T, Wang X (2021) The effectiveness of the problem-based learning in medical cell biology education: a systematic meta-analysis. Medicine 100(39):e27402
Yaduvanshi S, Singh S (2019) Fostering achievement of low-, average-, and high-achievers students in biology through structured cooperative learning (STAD method). Educ Res Int 2019:1–10. https://doi.org/10.1155/2019/1462179

Author contributions
Conception or design of the work: ZA, JB, MES; data collection: MS; data analysis and interpretation: MES; drafting the article: MES, ZA; critical revision of the article: MES, ZA, JB; final approval of the version to be published: MES, ZA, JB; and NA-L.

Funding
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interests
The authors declare no competing interests.

Ethical approval
Ethical approval was not required as the study did not involve human participants.

Informed consent
The article does not contain any studies with human participants performed by any of the authors.

Additional information
Correspondence and requests for materials should be addressed to Jolly Bhadra.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.