MODESTUM EURASIA Journal of Mathematics, Science and Technology Education, 2026, 22(10), em2929
ISSN:1305-8223 (online)
OPEN ACCESS Research Paper https://doi.org/10.29333/ejmste/19401

Gamification in secondary school physics: An exploratory quasi-experimental study of conceptual understanding in energy and momentum

Marija Gaurina [1*], Ivana Weber [1], Jerneja Pavlin [2]

[1] Faculty of Science, University of Split, Split, CROATIA
[2] Faculty of Education, University of Ljubljana, Ljubljana, SLOVENIA

Received 09 June 2026 ▪ Accepted 04 September 2026

Abstract

Energy and momentum are fundamental concepts in secondary-school physics, yet students often continue to experience conceptual difficulties even after formal instruction. This study explored differences in students' conceptual understanding between gamified and non-gamified instructional conditions. A classroom-based quasi-experimental pre-/post-test design was implemented with 63 students aged 15-16 from three intact first-year gymnasium classes in Croatia. One class received gamified instruction, while two classes received the same curricular content, learning activities, instructional time, teacher guidance, and feedback without the game-design elements. Students' conceptual understanding was assessed using a Croatian translation of the energy and momentum conceptual survey (EMCS) administered before and after the eight-week instructional unit. Student-level analyses showed a larger mean raw gain in the gamified condition than in the non-gamified condition. A baseline-adjusted analysis using HC3 heteroscedasticity-consistent standard errors showed the same general pattern. The gamified condition also demonstrated a higher class-average normalized gain, although both conditions remained within Hake's (1998) low-gain range. Several EMCS items remained challenging after instruction in both conditions. Because only one intact class received the gamified intervention, instructional condition could not be separated fully from class membership. The findings should therefore be interpreted as context-specific and exploratory rather than as definitive evidence of a causal effect of gamification.

Keywords: conceptual understanding, energy and momentum, gamification, physics education, secondary school

INTRODUCTION

A sound understanding of energy and momentum is fundamental to the study of mechanics in secondary-school physics. In the Croatian curriculum for the first year of gymnasium, students are expected not only to solve numerical problems but also to develop a conceptual understanding of work, energy transformations, and momentum conservation, enabling them to relate physical principles to real-world phenomena (Ministarstvo znanosti i obrazovanja Republike Hrvatske, 2019).

Despite these curricular goals, research has consistently shown that many students complete physics instruction with fragmented or scientifically inaccurate understandings of energy and momentum. Difficulties in interpreting and applying concepts such as work, energy, impulse, and momentum often persist even after formal instruction (Brundage et al., 2023; Singh & Rosengrant, 2003). Similar challenges have been reported across scientific disciplines, suggesting that students may construct context-dependent rather than coherent conceptions of energy (Opitz et al., 2017).

Conceptual inventories such as the energy and momentum conceptual survey (EMCS) (Singh &

This article is part of the first author's PhD thesis from faculty of science at the University of Split in Croatia.

✉ mgaurina@pmfst.hr **(*Correspondence)** ✉ ivana@pmfst.hr ✉ jerneja.pavlin@pef.uni-lj.si

Contribution to the literature

- This study contributes context-specific evidence on differences in conceptual understanding associated with gamified and non-gamified instruction in Croatian secondary-school physics.
- It examines gamification as an additional game-design layer within otherwise equivalent instruction, rather than as a comparison between different curricular content or teaching activities.
- It combines overall, class-specific, and item-level findings on energy and momentum concepts to identify both the more favorable pattern of conceptual understanding observed in the gamified class and the conceptual difficulties that remained after instruction.

Rosengrant, 2003) have further demonstrated the persistence of these difficulties. Even students who have completed advanced mechanics courses may experience problems when reasoning about fundamental energy and momentum concepts, particularly when they are required to apply their knowledge in unfamiliar situations (Brundage et al., 2023; Maries et al., 2023). Successful completion of conventional instruction therefore does not necessarily ensure durable conceptual understanding.

These challenges have stimulated interest in instructional approaches that actively involve students in learning and support sustained engagement with demanding conceptual content. One such approach is gamification, commonly defined as the use of game-design elements in non-game contexts (Deterding et al., 2011). In education, gamification may incorporate points, levels, badges, progression structures, collaborative challenges, and competitive elements into otherwise regular learning activities.

Gamification has become an increasingly prominent topic in educational research because it offers alternative ways of structuring participation, progression, and classroom engagement (Murillo Zamorano et al., 2021). A systematic review of gamification in primary and secondary education identified generally favorable outcomes but also emphasized substantial variation in implementation, research design, and measured outcomes (Vrcelj et al., 2023). Similarly, Dichev and Dicheva (2017) concluded that the success of gamification depends less on the mere presence of game elements than on the quality of the underlying instructional design. A meta-analysis by Sailer and Homner (2020) reported generally positive cognitive, motivational, and behavioral outcomes while also documenting considerable heterogeneity across educational contexts and intervention designs.

Gamification is therefore unlikely to lead to universal improvements. Its educational value appears to depend on the extent to which game-design elements are aligned with learning objectives, meaningful activities, and students' motivational needs. Research in technology-supported learning environments has shown that gamification may be more beneficial when it is embedded in purposeful learning activities and supported by clear goals and active participation (Ibáñez et al., 2014). Experimental work has also indicated that individual game elements may contribute differently to students' concentration, engagement, enjoyment, and learning (Wang & Lieberoth, 2016). Motivational processes and students' perceptions of the instructional environment may consequently play an important role in shaping learning outcomes (Almisad & Aleidan, 2025; Dichev et al., 2022; Fernando & Premadasa, 2024; Ortiz Rojas et al., 2025).

Comparatively fewer studies have investigated gamification specifically in physics education. Gamification has been used to structure participation and increase engagement in undergraduate physics classrooms (Rose et al., 2016), while gamified and simulation-supported environments have been associated with more interactive learning experiences in mechanics (Forndran & Zacharias, 2019; Radnai et al., 2019; Richter & Kickmeier-Rust, 2025). However, the evidence base concerning conceptual understanding is less extensive than that concerning motivation or engagement. Conceptual understanding of work, energy, impulse, and momentum has also been found to be better supported by instructional approaches that encourage students to examine relationships among concepts rather than rely on memorized procedures (Alonzo & Mistades, 2021). Digital platforms have also been used to assess students' understanding of energy and momentum and provide flexible forms of formative assessment (Yana et al., 2022). Nevertheless, students' conceptions may remain fragmented, context-dependent, and scientifically non-aligned after instruction (Brundage et al., 2023; Dega & Govender, 2016; Wu et al., 2023). These findings reinforce the view that game elements are most likely to support learning when they are integrated into a coherent pedagogical design that promotes conceptual reasoning and active participation (Richter & Kickmeier-Rust, 2025; Triantafyllou et al., 2025).

Within the Croatian educational context, teachers generally express positive attitudes towards the educational use of gamification, although many report limited opportunities for formal professional development and insufficient classroom resources for its implementation (Gaurina & Pavlin, 2025). Despite this growing interest, evidence from Croatian secondary-school physics classrooms remains limited.

The present study addresses this gap by examining differences in students' conceptual understanding under gamified and non-gamified instructional conditions using a Croatian translation of the EMCS. Both instructional conditions covered the same curricular content and underlying learning activities, while the gamified condition incorporated an additional layer of game-design elements. In addition to examining overall changes in conceptual understanding, the study identifies EMCS items and associated conceptual areas that remained challenging after instruction.

Accordingly, the study was guided by the following research questions (RQs):

RQ1. How do gains in conceptual understanding of energy and momentum differ between gamified and non-gamified instructional conditions?

RQ2. Which EMCS items and associated conceptual areas remain most challenging after instruction, and how do these patterns differ between the gamified and non-gamified instructional conditions?

MATERIALS AND METHODS

The study employed a classroom-based quasi-experimental pre-/post-test design to examine differences in secondary-school students' conceptual understanding of energy and momentum under gamified and non-gamified instructional conditions. The intervention was implemented in regular first-year gymnasium physics classes in Croatia and was aligned with the national physics curriculum (Ministarstvo znanosti i obrazovanja Republike Hrvatske, 2019). Both conditions covered the same curricular content, learning outcomes, instructional activities, and amount of instructional time. The main difference was that, on one condition, the learning activities were additionally structured through game-design elements.

Participants

Initially, 69 first-year gymnasium students from three intact classes participated in the study during the 2021/2022 school year. Six students were excluded from the final analyses because they did not complete both assessments and/or did not attend the instructional unit. The final analytic sample therefore comprised 63 students. Two classes received non-gamified instruction and formed the non-gamified condition ($N = 42$), while one class received a gamified version of the same lessons and formed the gamified condition ($N = 21$). Baseline EMCS performance was assessed before instruction and formally compared between the two instructional conditions. All classes were taught by the same instructor, which reduced between-teacher variability in curriculum coverage and general teaching approach. However, because the teacher was aware of the instructional condition implemented in each class, expectancy effects cannot be entirely excluded.

The study was conducted in three pre-existing first-year gymnasium classes selected by the school administration for participation because they followed the same curriculum and were considered broadly comparable in their educational background. After these three intact classes had been identified for inclusion in the study, the classes—not individual students—were randomly allocated to the instructional conditions using a computer-based randomization application. One intact class was allocated to the gamified condition and two intact classes to the non-gamified condition. Students remained in their existing classes throughout the study and were not individually randomized or reassigned. Because randomization involved only three intact classes in a 1:2 allocation, it could not be expected to ensure balance in class-level characteristics or baseline performance. Indeed, although the pooled gamified and non-gamified conditions had nearly identical mean pre-test EMCS scores, the two non-gamified classes differed in their baseline means ($M = 3.10$, $SD = 0.77$ and $M = 5.05$, $SD = 0.92$), with the gamified class falling between them ($M = 4.05$, $SD = 1.24$). Instructional condition was therefore partially confounded with class membership, and the small number of participating classes did not permit reliable separation of instructional effects from class-specific characteristics. Although all three classes followed the same curriculum and were taught by the same teacher, unmeasured class-level factors, including classroom dynamics, peer composition, or timetable-related circumstances, cannot be excluded as alternative explanations for the observed differences.

Before data collection, students and their parents or legal guardians were informed about the purpose and procedures of the study. Written informed consent was obtained from parents or legal guardians, and the participating students provided written assent before participation. Participation was voluntary, and students were informed that they could withdraw without any adverse consequences. Anonymity, confidentiality, and data protection were ensured throughout the study. Participation in the research and students' EMCS results had no influence on their school grades. The study was conducted in accordance with the applicable institutional ethical principles and professional standards.

Instrument, Procedures, and Data Collection

Data were collected using the EMCS, originally developed by Singh and Rosengrant (2003). The EMCS is a 25-item multiple-choice diagnostic instrument designed to assess students' conceptual understanding of fundamental principles related to work, energy, impulse, and momentum. Subsequent research has also explored the development of EMCS-based four-tier diagnostic questions to examine students' understanding of energy and momentum concepts in greater detail (Afif et al., 2017). For clarity and consistency, the items are labelled Q1, Q2, Q3, and so forth throughout the paper.

Because the EMCS was originally developed in English, the first author translated the instrument into Croatian, and a qualified English-language teacher subsequently back-translated it and compared the back-translation with the original. The comparison confirmed agreement in meaning and terminology, and no discrepancies requiring revision were identified. The Croatian version was then piloted with 22 students from a Croatian vocational secondary school who were not included in the present study. Students considered the clarity of the instructions, item wording, the comprehensibility of the physics-related terms, and response options. All students indicated that the instructions and items were understandable, and no ambiguities were identified.

Therefore, the final Croatian version of the EMCS retained the original 25-item structure, conceptual content, response options, scoring procedure, and intended conceptual focus. The content represented in the EMCS corresponded to the topics covered during the first-year gymnasium instructional unit, including work, kinetic and potential energy, energy transformations, conservation of mechanical energy, impulse, momentum, and conservation of momentum. Nevertheless, linguistic equivalence and curricular alignment do not necessarily indicate that the instrument is fully suitable for use with students aged 15-16. The Croatian EMCS was therefore treated as a translated research instrument rather than as an independently validated Croatian version. In their original development study, Singh and Rosengrant (2003) reported a Cronbach's α of 0.75 for a sample of 1,170 students and 0.68 for a sample of 186 students enrolled in algebra-based classes.

KR-20 was calculated separately for the pre-test and post-test responses of the 63 students. The coefficients were -1.038 for the pre-test and -0.243 for the post-test. These negative coefficients do not provide evidence of satisfactory internal consistency and indicate that, in this sample and these administrations, the 25 dichotomously scored items did not function as a homogeneous set. The coefficients were therefore not interpreted as evidence of satisfactory score reliability for the Croatian administration of the EMCS.

The EMCS samples several related and partly overlapping aspects of reasoning about energy and momentum. Internal-consistency estimates should therefore be interpreted in relation to the instrument's dimensionality, content coverage, and intended use and should not, on their own, be treated as evidence of validity or overall measurement quality (Sijtsma, 2009; Taber, 2018). Nevertheless, the negative KR-20 coefficients obtained in the present sample represent an important measurement limitation and reinforce the need for a more comprehensive psychometric examination of the Croatian EMCS in a larger secondary-school sample.

The EMCS was administered in paper-and-pencil form before and after an eight-week instructional unit. Both instructional conditions covered the same curricular content, learning outcomes, sequence of topics, instructional time, learning materials, problem-solving activities, quizzes, collaborative activities, teacher guidance, and feedback. The same teacher taught all three classes using comparable pacing and the same underlying instructional activities.

The only systematic difference between conditions was the addition of game-design elements in the gamified class. In that class, the existing learning activities were framed through points, badges, level progression, competitive elements, and collaborative activities presented as team challenges. Students in the non-gamified classes completed the same underlying activities without these game-design elements. Gamification therefore functioned as an additional layer within otherwise equivalent instruction rather than as a separate pedagogical approach. The game-design elements were aligned with curricular objectives and with the motivational principles discussed by Dichev and Dicheva (2017) and Hamari et al. (2014), particularly those concerning competition, progression, and sustained participation.

Data Analysis

Data were analyzed using descriptive and inferential statistical procedures (Pallant, 2020). Raw data were entered into Microsoft Excel and analyzed using IBM SPSS Statistics, version 26. Statistical significance was evaluated at the .05 level for the student-level inferential analyses. However, all *p* values and confidence intervals were interpreted with caution, given that students were nested within only three intact classes.

Item-level response patterns were examined to identify which EMCS items appeared to show relatively greater gains and which remained challenging under each instructional condition. Because students' reasoning for their answer choices on the multiple-choice items was not collected, these response patterns were interpreted as indicators of relative conceptual difficulty rather than as direct evidence of specific alternative conceptions. Items were also mapped to one or more conceptual domains to provide a descriptive overview of the content represented in the EMCS. Because several items address overlapping concepts, the mapping was intentionally non-exclusive. The resulting conceptual groupings were used descriptively and were not treated as statistically independent subscales or domain-level measures. **Figure 1** illustrates the non-exclusive mapping of items to conceptual domains.

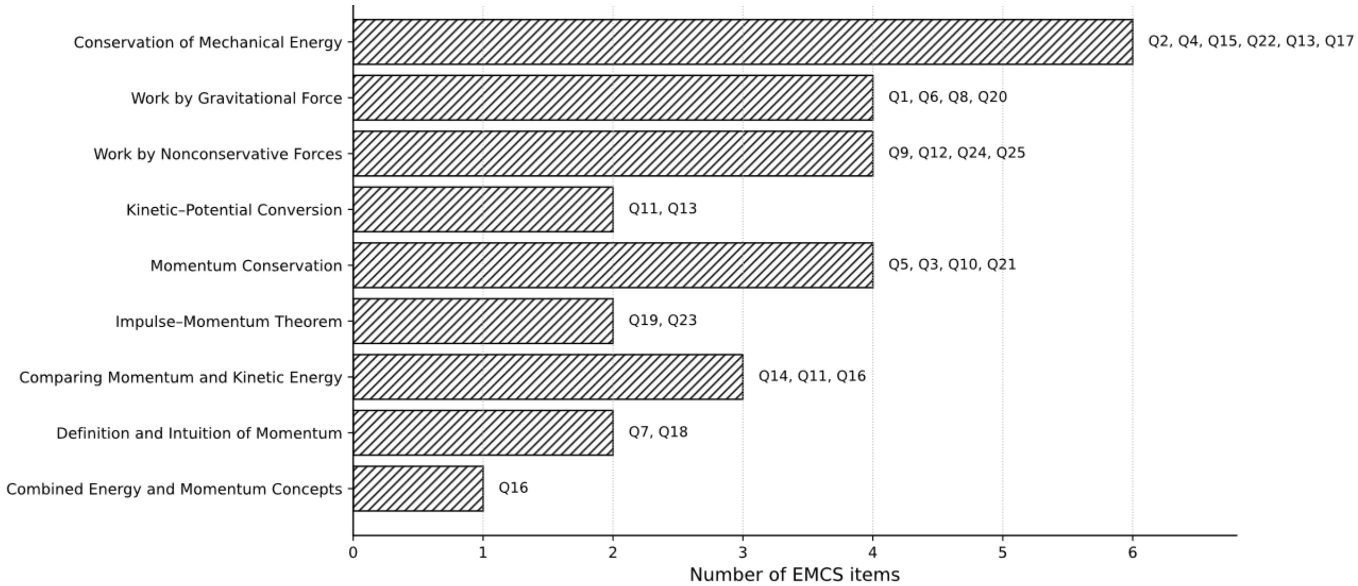

Figure 1. Non-exclusive mapping of EMCS items across conceptual domains in energy and momentum (the authors' own elaboration)

Students' understanding was examined using raw EMCS gain scores and normalized gain. Raw gain was calculated as the post-test score minus the pre-test score. Individual normalized gain (g_i) was calculated for each student using Eq (1):

$$g_i = \frac{(post\%-pre\%)}{(100\%-pre\%)}. \quad (1)$$

The means and standard deviations of the individual normalized gains are reported as M_g and SD_g. To permit comparison with Hake's (1998) original convention, a class-average normalized gain, $\langle g \rangle$, was also calculated from the mean pre-test and post-test percentages for each instructional condition. Item-level normalized gains were calculated descriptively from the condition-level percentages of correct responses on each item. Hake's (1998) conventional categories were applied only to the class-average $\langle g \rangle$ values and were not treated as validated thresholds for individual students, individual items, or the non-exclusive conceptual groupings. Because the EMCS is a diagnostic conceptual inventory rather than a criterion-referenced achievement test, percentage scores were interpreted as performance on the instrument and were not mapped onto Croatian school-grading or state-examination thresholds.

Within-condition changes were examined using paired-samples *t*-tests. For the paired-samples analyses, the distribution of within-student difference scores (post-test minus pre-test) was examined separately for each instructional condition using skewness, excess kurtosis, the Shapiro-Wilk test, and inspection for extreme outliers. Because the difference-score distributions showed departures from normality, Wilcoxon signed-rank tests were additionally conducted as non-parametric sensitivity analyses. Raw gain scores were compared between instructional conditions using Welch's independent-samples *t*-test because the groups differed in size and the gain-score variances were unequal. Welch's test provides better control of type I error than student's *t*-test when the assumption of equal variances is violated (Delacre et al., 2017; Ruxton, 2016). The standardized between-condition difference was reported using Hedges' g with a 95% confidence interval (CI).

Baseline EMCS scores were also compared using Welch's independent-samples *t*-test and Hedges' g. As an additional sensitivity analysis, post-test EMCS performance was examined using a baseline-adjusted model equivalent to ANCOVA, with instructional condition as the predictor and pre-test EMCS score as the covariate. Linearity was examined using scatterplots and an additional quadratic pre-test term. Homogeneity of regression slopes was evaluated through the pre-test-by-condition interaction. Residual normality and heteroscedasticity were examined using diagnostic plots, the Shapiro-Wilk test, and the Breusch-Pagan test. Cook's distance was used to examine potentially influential observations. Because the residual diagnostics indicated heteroscedasticity, coefficients and statistical tests were evaluated using HC3 heteroscedasticity-consistent standard errors.

The class-level allocation and inclusion of only three intact classes did not permit reliable multilevel modelling or cluster-adjusted estimation. The primary inferential analyses were therefore conducted at the student level and interpreted as exploratory estimates of differences associated with instructional condition. Additional class-specific analyses were conducted to determine whether the observed pooled difference depended on the performance of either non-gamified class. The two non-gamified classes were first compared with one another, after which the gamified class was compared separately with each non-gamified class using

Table 1. Class-specific descriptive EMCS results

Class	Instructional condition	N	Pre-test *M (SD)*	Post-test *M (SD)*	Raw gain *M (SD)*	Individual normalized gain M_g (SD_g)
Gamified class	Gamified	21	4.05 (1.24)	9.81 (1.78)	5.76 (1.81)	0.274 (0.082)
Non-gamified class 1	Non-gamified	21	3.10 (0.77)	5.86 (0.65)	2.76 (0.54)	0.126 (0.023)
Non-gamified class 2	Non-gamified	21	5.05 (0.92)	7.71 (0.78)	2.67 (0.48)	0.133 (0.022)

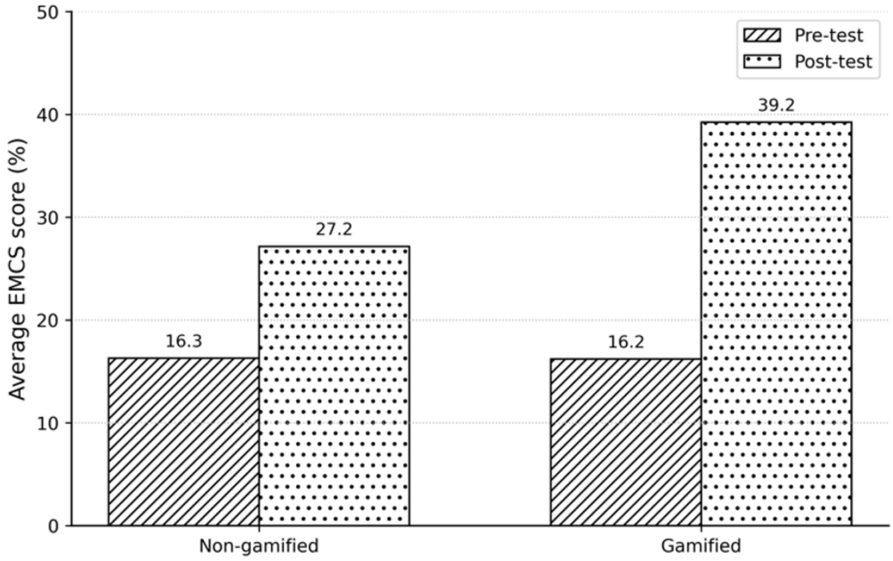

Figure 2. Average pre- and post-test EMCS scores by instructional condition (the authors' own elaboration)

Welch's t-tests, Hedges' g, and baseline-adjusted models with HC3 standard errors. These analyses were treated as exploratory robustness checks and could not separate instructional condition from characteristics specific to the single gamified class.

RESULTS

Results are presented according to RQs.

RQ1. Differences in Conceptual Gains in Energy and Momentum Between Gamified and Non-Gamified Instruction

Before examining changes over time, baseline EMCS performance was compared between the two instructional conditions. On the 25-item instrument, with one point awarded for each correct response (possible score range: 0-25), mean pre-test scores were nearly identical in the gamified condition (M = 4.05, SD = 1.24) and the pooled non-gamified condition (M = 4.07, SD = 1.30). Welch's t-test indicated no statistically significant pooled baseline difference, t (41.60) = -0.07, p = .944. The standardized difference was negligible, Hedges' g = -0.02, 95% CI [-0.54, 0.50]. Although the pooled conditions had similar mean baseline scores, this similarity masked class-level variation between the two non-gamified classes and does not establish equivalence in unmeasured characteristics.

Descriptive results were therefore also examined separately for each participating class. As shown in **Table 1**, the two non-gamified classes differed in baseline EMCS performance but demonstrated nearly identical raw and individual normalized gains. The gamified class showed a larger mean gain than either non-gamified class.

The two non-gamified classes had mean baseline scores of 3.10 and 5.05, respectively, but their mean raw gains were 2.76 and 2.67 points. Their mean individual normalized gains were similarly close, M_g = .126 and M_g = .133. The pooled non-gamified result was therefore not descriptively driven by one comparison class, although the class-level dependency remained.

Figure 2 presents the mean pre-test and post-test EMCS percentages for the two instructional conditions. Both conditions began with approximately 16% correct responses. Following instruction, mean performance increased to 27.2% in the non-gamified condition (M = 6.79, SD = 1.18) and to 39.2% in the gamified condition (M = 9.81, SD = 1.78). Absolute performance therefore remained modest in both conditions, although the observed increase was larger in the gamified condition. Because the EMCS was originally developed for undergraduate populations, the percentages are interpreted as indicators of performance on this instrument rather than as school-grade equivalents.

Before conducting the paired-samples t-tests, the distributions of the within-student difference scores (post-test minus pre-test) were examined separately for the two instructional conditions. In the gamified condition, skewness was .336 and excess kurtosis was -1.293, whereas in the non-gamified condition, skewness was -.383 and excess kurtosis was -.556. Shapiro-Wilk

Table 2. Exploratory comparisons of the gamified class with each non-gamified class on EMCS raw gain and baseline-adjusted post-test scores

Comparison	Welch's *t*	*df*	*p*	Hedges' *g*	Adjusted *B*	HC3 *SE*	95% CI	*p*
Gamified class vs non-gamified class 1	7.26	23.50	< .001	2.20	3.47	0.44	[2.59, 4.36]	< .001
Gamified class vs non-gamified class 2	7.56	22.82	< .001	2.29	2.65	0.41	[1.81, 3.48]	< .001

Note. Welch's t tests and Hedges' g were calculated for individual raw gain scores; The baseline-adjusted models included post-test EMCS score as the outcome, instructional condition as the predictor, and pre-test EMCS score as the covariate; Adjusted condition coefficients are reported with HC3 heteroscedasticity-consistent standard errors and 95% confidence intervals; & All analyses were conducted at the student level and are interpreted as exploratory robustness checks rather than as cluster-adjusted estimates

tests indicated departures from normality in both the gamified condition, $W = .894$, $p = .027$, and the non-gamified condition, $W = .668$, $p < .001$. No extreme outliers were identified in either condition. Because the difference scores were discrete and restricted in range, Wilcoxon signed-rank tests were additionally conducted as sensitivity analyses and showed significant pre-post improvement in both conditions (both $p < .001$).

Paired-samples *t*-tests likewise showed statistically significant pre-post improvement in both conditions. In the gamified condition, the mean score increased from $M = 4.05$ ($SD = 1.24$) to $M = 9.81$ ($SD = 1.78$), $t\,(20) = -14.56$, $p < .001$. In the non-gamified condition, the mean score increased from $M = 4.07$ ($SD = 1.30$) to $M = 6.79$ ($SD = 1.18$), $t\,(41) = -34.64$, $p < .001$.

Raw gain scores were calculated for each student as the difference between post-test and pre-test performance. The gamified condition demonstrated a mean raw gain of 5.76 points ($SD = 1.81$), compared with 2.71 points ($SD = 0.51$) in the non-gamified condition. Levene's test indicated unequal gain-score variances, $F\,(1, 61) = 62.57$, $p < .001$. The between-condition comparison was therefore conducted using Welch's *t*-test. The observed mean raw gain was significantly larger in the gamified condition, $t\,(21.58) = 7.55$, $p < .001$. The mean difference was 3.05 points, 95% CI [2.21, 3.89], with Hedges' $g = 2.69$, 95% CI [1.99, 3.39]. The very large standardized difference should be interpreted cautiously because its magnitude was influenced by the comparatively low variability of gain scores, particularly in the non-gamified condition ($SD = 0.51$), and should not be interpreted as indicating a correspondingly large absolute understanding gain.

The mean individual normalized gain was $M_g = .274$ ($SD_g = .082$) in the gamified condition and $M_g = .130$ ($SD_g = .023$) in the non-gamified condition. The corresponding class-average normalized gains, calculated from the condition-level pre-test and post-test means, were $\langle g \rangle = .275$ and $\langle g \rangle = .130$, respectively. Both values fall within Hake's (1998) low-gain range. The higher value in the gamified condition therefore represents a more favorable relative pattern of improvement, but not a high absolute gain.

A baseline-adjusted sensitivity analysis was conducted with post-test EMCS performance as the outcome, instructional condition as the predictor, and pre-test EMCS performance as the covariate. The pre-test-by-condition interaction was not statistically significant when evaluated using HC3 standard errors, $B = -0.380$, HC3 $SE = 0.245$, $t\,(59) = -1.55$, $p = .127$. This result provided no clear evidence that the relationship between pre-test and post-test performance differed between the two instructional conditions. The addition of a quadratic pre-test term did not improve the model, $F\,(1, 59) = 0.56$, $p = .456$, providing no evidence of substantial non-linearity.

Residual diagnostics indicated heteroscedasticity, Breusch-Pagan $\chi^2\,(2) = 23.66$, $p < .001$, and a modest departure from normality, Shapiro-Wilk $W = .957$, $p = .028$. The final estimates were therefore evaluated using HC3 heteroscedasticity-consistent standard errors. After adjustment for pre-test performance, instructional condition remained significantly associated with post-test EMCS score, $B = 3.041$, HC3 $SE = 0.394$, $t\,(60) = 7.71$, $p < .001$, 95% CI [2.25, 3.83].

The maximum Cook's distance was 0.159, and no individual observation exerted excessive influence on the model. Leave-one-out estimates of the adjusted condition difference ranged from 2.88 to 3.18 points and remained statistically significant. The baseline-adjusted analysis therefore showed the same general pattern as the raw gain analysis. However, it represents a complementary analysis of the same students rather than independent confirmation of an instructional effect, and possible class-level confounding remains.

Additional class-specific analyses were conducted to explore whether the pooled between-condition difference may have been influenced by the performance of either non-gamified class. Despite their different baseline scores, the two non-gamified classes had nearly identical raw gains, 2.76 ($SD = 0.54$) and 2.67 ($SD = 0.48$), Welch's $t\,(39.46) = 0.60$, $p = .550$. Their individual normalized gains were also similar, $M_g = .126$ ($SD_g = .023$) and $M_g = .133$ ($SD_g = .022$), Welch's $t\,(39.92) = -1.09$, $p = .282$. The gamified class was then compared separately with each non-gamified class using Welch's *t*-tests of raw gain, Hedges' *g*, and baseline-adjusted models with HC3 standard errors. The results are presented in **Table 2**.

The gamified class demonstrated larger raw gains than each non-gamified class, with large observed standardized differences. The baseline-adjusted models with HC3 standard errors showed the same general pattern in both pairwise comparisons. These analyses

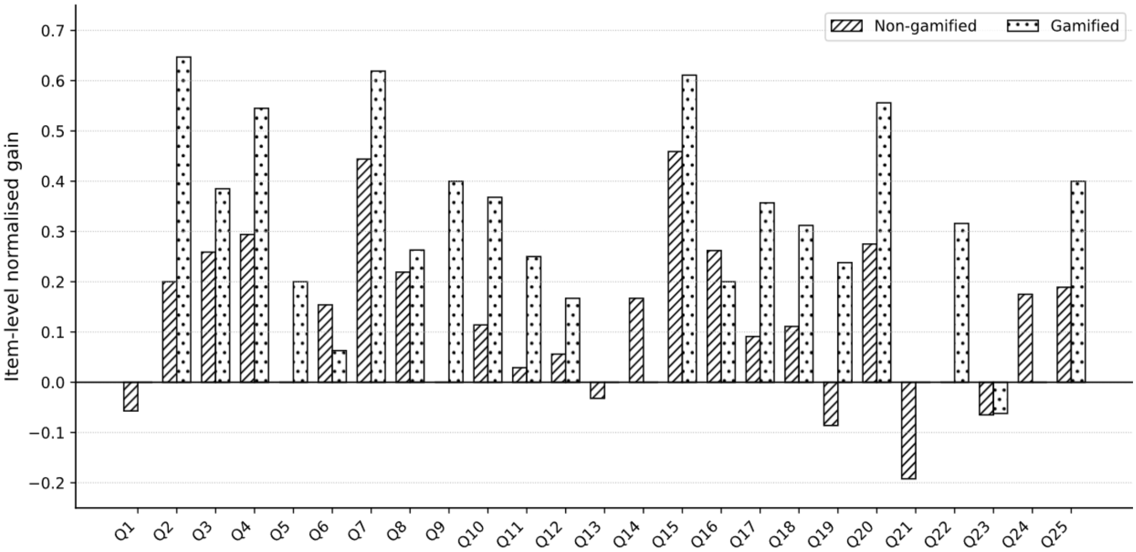

Figure 3. Normalized knowledge gains for individual EMCS items (Q1-Q25) in the gamified and non-gamified conditions (the authors' own elaboration)

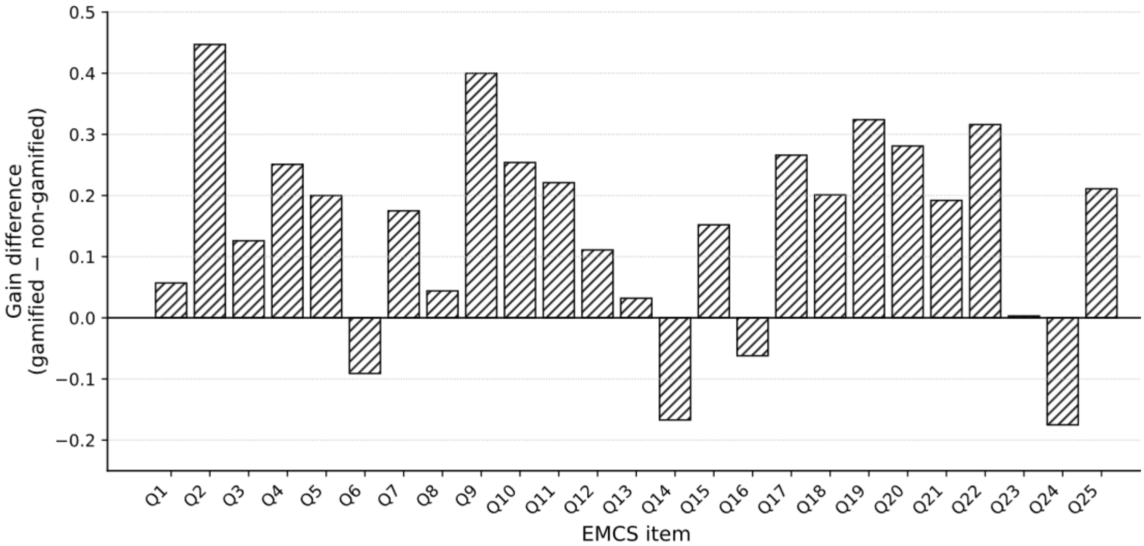

Figure 4. Differences in normalized knowledge gains between the gamified and non-gamified conditions across EMCS items (the authors' own elaboration)

indicate that the pooled difference was not descriptively attributable solely to one of the two non-gamified classes. Nevertheless, because only one intact class received the gamified intervention, the comparisons cannot distinguish instructional condition from characteristics specific to that class. They should therefore be interpreted as exploratory checks of the consistency of the observed pattern rather than as independent or cluster-adjusted estimates of a gamification effect.

Item-level normalized gains were examined to describe how the observed improvement was distributed across the EMCS. As shown in Figure 3, the gamified condition tended to show higher gains on most items, including gains greater than .50 for Q2, Q4, Q7, Q15, and Q20. The non-gamified condition showed fewer items with gains of a comparable magnitude. These item-level values are descriptive and should not be interpreted using Hake's (1998) class-average gain categories.

Additional insight can be gained from examining differences in item-level normalized gains between the two conditions. These values, calculated by subtracting the non-gamified value from the corresponding gamified value, are presented in Figure 4. Positive values indicate a more favorable pattern in the gamified condition, whereas negative values indicate a more favorable pattern in the non-gamified condition. Most items showed positive differences, with particularly large differences for Q2, Q3, Q4, Q7, Q10, Q15, Q20, and Q25. Five items—Q6, Q14, Q16, Q23, and Q24—showed negative or near-zero differences. The relative advantage observed for the gamified condition was therefore not uniform across the instrument.

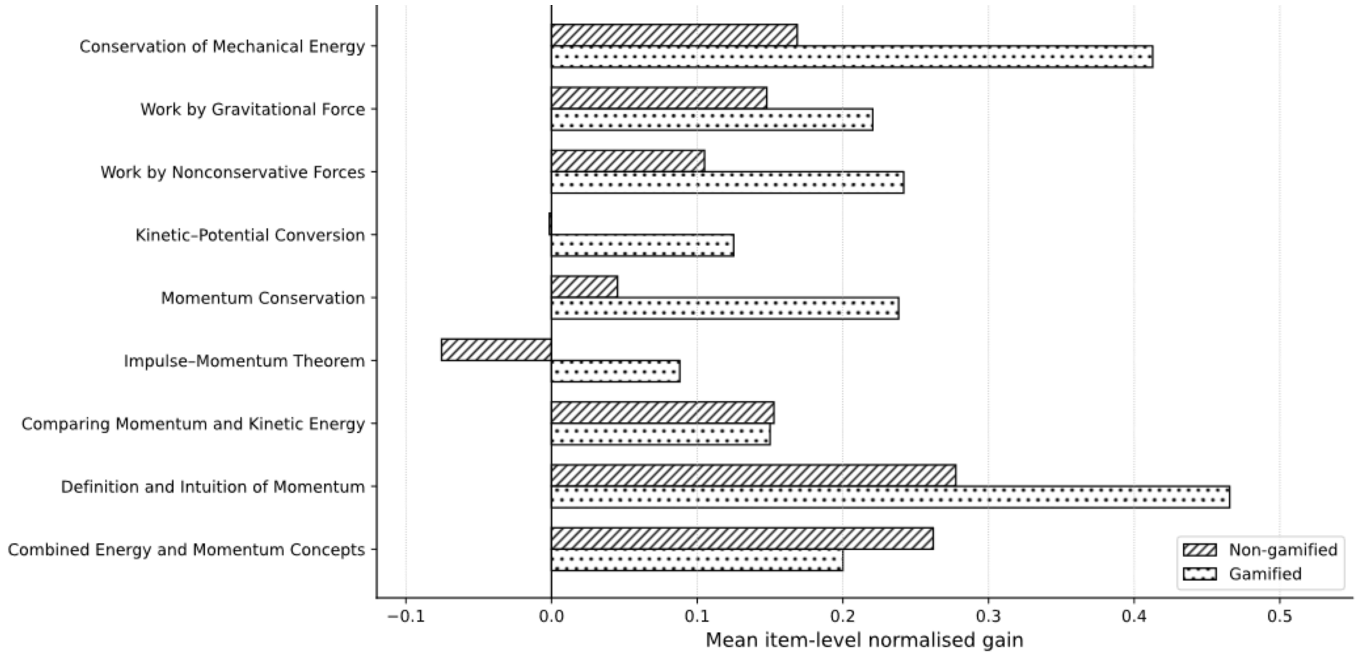

Figure 5. Descriptive comparison of mean item-level normalized gains across non-exclusive conceptual groupings in the gamified and non-gamified conditions (the authors' own elaboration)

RQ2. Comparison of EMCS Items and Conceptual Areas Remaining Challenging After Instruction

An item-level analysis identified several EMCS items that remained challenging after instruction, as indicated by low post-test performance accompanied by limited or negative pre-post change. A fixed normalized-gain threshold was not applied because item-level gain values can be sensitive to initial performance and small changes in the number of correct responses.

In the gamified condition, particularly limited improvement was observed for Q1, Q13, Q23, and Q24, which address gravitational work, conservation of mechanical energy, the impulse-momentum theorem, and work by nonconservative forces, respectively. Post-test accuracy was 5% for Q1, 33% for Q13, and 19% for both Q23 and Q24. Q23 showed a small negative normalized gain, g = −.062, while Q1, Q13, and Q24 showed no improvement. Similar low or negative patterns were observed for several of these items in the non-gamified condition, indicating that the corresponding content remained challenging under both instructional approaches.

Because students' justifications for their answer choices on the multiple-choice EMCS were not collected, these observed response patterns should not be interpreted as direct evidence of particular persistent alternative conceptions. Instead, they identify content areas that may require more targeted instructional interventions, including conceptual conflict, scaffolding, multiple representations, and opportunities for students to articulate, compare and evaluate their reasoning.

Figure 5 provides a descriptive comparison of mean item-level normalized gains across the non-exclusive conceptual groupings. Because several EMCS items contribute to more than one grouping, these values should not be interpreted as independent subscale scores or subjected to inferential comparisons. The gamified condition showed higher mean gains in seven of the nine conceptual groupings. The largest mean gains in this condition were observed for definition and intuition of momentum (g = .466) and conservation of mechanical energy (g = .413). More modest gains were observed for work by nonconservative forces (g = .242), momentum conservation (g = .238), work by gravitational force (g = .220), and combined energy and momentum concepts (g = .200). The smallest positive values in the gamified condition were found for kinetic-potential conversion (g = .125) and the impulse-momentum theorem (g = .088).

In the non-gamified condition, the largest mean gains were observed for definition and intuition of momentum (g = .278) and combined energy and momentum concepts (g = .262). Particularly limited improvement was evident for momentum conservation (g = .045), while kinetic-potential conversion (g = -.002) and the impulse-momentum theorem (g = -.075) showed near-zero or negative mean gains. The two conditions showed almost identical values for comparing momentum and kinetic energy (g = .150 in the gamified condition and g = .152 in the non-gamified condition), whereas the non-gamified condition showed a somewhat higher value for combined energy and momentum concepts (g = .262) than the gamified condition (g = .200). These category-level findings are descriptive and are used to identify broad patterns across related content areas; the individual-item results remain the primary basis for interpreting which EMCS content remained challenging after instruction.

DISCUSSION

The present study examined differences in students' conceptual understanding of energy and momentum following gamified and non-gamified instruction in secondary-school physics. Across complementary analyses of the same dataset, the observed patterns generally favored the gamified class over the two non-gamified classes. The raw gain, normalized gain, baseline-adjusted, and class-specific analyses revealed broadly similar patterns. However, because these analyses were based on the same three intact classes, they should not be interpreted as independent confirmations of an instructional effect.

The magnitude of the observed difference was evident in both the raw and normalized gain results. The mean individual normalized gain was M_g = .274 in the gamified condition and M_g = .130 in the non-gamified condition. The corresponding class-average values were ⟨*g*⟩ = .275 and ⟨*g*⟩ = .130. Both fall within Hake's (1998) low-gain range. The larger value in the gamified condition therefore reflects a more favorable relative pattern of improvement, but it should not be described as a high or substantial absolute gain. Similarly, the very large Hedges' *g* for the raw-gain comparison should be interpreted in light of the low gain-score variability, particularly in the non-gamified condition, rather than as evidence of a comparably large absolute understanding gain.

The class-specific findings provide useful context for interpreting the pooled result. Although the two non-gamified classes differed substantially in baseline EMCS performance, they achieved almost identical raw and individual normalized gains. The gamified class demonstrated larger gains than each of them when the comparisons were conducted separately. This pattern indicates that the pooled result was not descriptively attributable solely to unusually weak improvement in one non-gamified class. It does not, however, resolve the central design limitation: only one intact class received the gamified intervention, and instructional condition cannot be separated fully from class membership. The observed differences should therefore be interpreted as context-specific associations within the participating classrooms rather than as definitive evidence of a causal effect of gamification.

Novelty and teacher-expectancy effects may also have contributed to the findings. The game-design elements differed from students' usual classroom experience, and the teacher was necessarily aware of the condition being implemented. The novelty of points, badges, levels, competitive framing, and team-based challenges may have temporarily increased attention or enthusiasm, while teacher awareness may have affected classroom interaction, reflecting the broader literature on teacher-expectancy effects, in which expectations can shape instructional behaviors and student performance (Rosenthal & Jacobson, 1968). These possibilities were not measured directly. Moreover, the eight-week duration does not allow short-term novelty responses to be distinguished from more sustained learning benefits. Replication across multiple teachers, classrooms, and longer implementation periods is therefore needed.

The relatively low absolute EMCS scores in both conditions also require careful interpretation. Mean post-test performance remained below 40% even in the gamified condition, indicating that substantial conceptual difficulties persisted. The EMCS was developed for undergraduate populations and assesses conceptual reasoning rather than routine procedural performance. The very low pre-test performance, approximately 16% correct, may partly reflect the difficulty of an instrument originally developed for undergraduate students when administered to 15-16-year-old secondary-school students before instruction on the relevant unit. Moreover, the 20% value expected from purely random responding to five-option multiple-choice items should not be interpreted as a performance threshold: students may systematically select conceptually plausible distractors rather than choose among the response options with equal probability. Because students' response processes and reasons for selecting particular distractors were not examined, however, the present data cannot determine why the mean pre-test performance fell below this theoretical random-response expectation.

Low scores should therefore not be interpreted simply as an absence of learning. Instead, the combination of measurable pre-post improvement and modest post-test performance suggests that students made progress while continuing to experience difficulties with several energy and momentum concepts. This pattern is consistent with previous research documenting persistent conceptual difficulties even after formal instruction (Brundage et al., 2023; Maries et al., 2023). Consistent with these findings, Singh and Rosengrant (2003), the developers of the instrument, reported an average post-test score of 49%. Nevertheless, when interpreted cautiously and at the descriptive level, the EMCS responses can provide information about energy and momentum items that remained challenging in this particular administration.

Item-level findings provide a more detailed view of the areas in which improvement occurred and those that remained difficult. Students in the gamified condition demonstrated larger gains on many items, but the pattern was not uniform. Q1, Q13, Q23, and Q24 remained particularly challenging. These questions require students to identify relevant physical principles, distinguish between closely related quantities, and apply conceptual reasoning within specific contexts rather than rely on direct recall or a single routine procedure. Previous EMCS research has similarly shown that students may struggle to apply energy and momentum principles in unfamiliar situations (Brundage et al., 2023; Singh & Rosengrant, 2003).

Because justifications for the selected answers on the EMCS items were not collected, the present findings cannot establish the specific reasoning patterns underlying incorrect responses. As noted by Singh and Rosengrant (2003), combining well-designed tests with interview data can provide deeper insight into students' conceptual reasoning. Nevertheless, the EMCS results may indicate that participation and task completion alone are insufficient for developing understanding in problems that require the integration of multiple concepts. The findings of the present study suggest a need for more explicit instructional support. This may include guided comparisons of alternative explanations, the use of multiple representations, the incorporation of conceptual conflict, and structured opportunities for students to justify their reasoning. Moreover, the results are consistent with previous research indicating that physics knowledge may remain fragmented and context-dependent even after formal instruction (Dega & Govender, 2016). In this respect, the findings suggest that gamification alone is unlikely to replace instructional strategies aimed specifically at fostering conceptual reconstruction.

The present design does, however, provide a relatively focused comparison because both conditions covered the same learning content, learning activities, quizzes, collaborative tasks, teacher guidance, and feedback. The principal difference was the game-design layer added to one class. Points, badges, progression structures, competitive framing, and team-based challenge presentation may have encouraged students to participate more consistently in demanding learning activities. Previous research suggests that gamification may support learning when game elements increase the perceived value of activities and encourage sustained participation (Dichev et al., 2022; Sailer & Homner, 2020). From an expectancy-value perspective, students are more likely to invest effort when an activity is perceived as both attainable and worthwhile (Eccles & Wigfield, 2002). However, because motivation and the proposed mediating processes were not measured, this explanation remains theoretically plausible rather than empirically demonstrated. The presented findings should not be interpreted as evidence that gamification is equally useful across all areas of physics learning. Its educational value is likely to depend on how successfully game elements are integrated with coherent activities that promote conceptual reasoning, discussion, and reflection. Previous research similarly emphasizes that game elements are more productive when they form part of a purposeful pedagogical design rather than function as isolated rewards (Hamari et al., 2014; Ortiz Rojas et al., 2025; Sailer & Homner, 2020). The present results are also comparable with findings from other structured instructional approaches in which overall conceptual understanding improved while difficulties remained in coordinating conceptual, symbolic, and graphical representations (Alonzo & Mistades, 2021).

The non-exclusive conceptual groupings should likewise be interpreted cautiously. Wu et al. (2023) showed that students' EMCS response patterns do not necessarily correspond closely to the expert-defined conceptual categories of the instrument. Conceptual understanding may extend across overlapping dimensions that do not align neatly with individual item classifications. This supports the decision to use the category-level results only as a descriptive overview and to place greater emphasis on individual-item patterns.

Taken together, the findings indicate that adding game-design elements to otherwise equivalent instruction was associated with a more favorable pattern of conceptual improvement in the participating classrooms. At the same time, the persistence of difficulty on several EMCS items shows that gamification alone is unlikely to produce comprehensive conceptual change.

CONCLUSIONS

The present study examined differences in students' conceptual understanding of energy and momentum following gamified and non-gamified instruction in Croatian secondary-school physics. Across complementary analyses of the same student-level data, the gamified class demonstrated a more favorable pattern of conceptual improvement than the two non-gamified classes. The two non-gamified classes achieved nearly identical gains despite differing baseline performance, while the gamified class showed larger gains than each comparison class. These findings indicate different observed learning patterns within the participating classrooms but do not provide an independent cluster-level estimate of a general gamification effect.

Despite the more favorable results in the gamified condition, substantial difficulties remained on several EMCS items addressing energy transformations, work, collisions, impulse, and momentum. Gamification should therefore not be viewed as a stand-alone solution for the conceptual difficulties reflected in students' responses. Its educational value is more likely to depend on how game-design elements are integrated with curriculum objectives and instructional strategies that explicitly support conceptual reasoning and reconstruction.

Several limitations should be considered. The study involved one gamified and two non-gamified intact classes taught by the same teacher. Instructional condition could not be separated fully from class membership, and the student-level analyses did not adjust for clustering. Possible novelty and teacher-expectancy influences were not measured, and the eight-week intervention does not establish whether the observed pattern would be sustained over time. Although the EMCS instrument underwent forward translation, back-translation, comparison with the original, and preliminary testing for comprehensibility, it has not been independently validated for Croatian secondary-school students. Moreover, the negative KR-20 coefficients obtained in the present sample did not provide interpretable evidence of satisfactory internal consistency, further underscoring the need for psychometric evaluation of the Croatian EMCS in a larger secondary-school sample. The findings should therefore be understood as score patterns from this administration rather than as validation evidence for the Croatian translation.

Within these limitations, the study contributes exploratory evidence from an educational context that remains underrepresented in research on gamification in physics education. It also illustrates the value of considering class-specific and item-level patterns alongside overall learning gains. Future research should replicate the intervention across multiple schools, teachers, and gamified classes, include longer follow-up periods, and combine conceptual inventories with measures of motivation and qualitative evidence of students' reasoning.

Author contributions: MG: conceptualization, data curation, formal analysis, writing – original draft, writing – review & editing, **IW:** conceptualization, formal analysis, writing – review & editing, **JP:** conceptualization, supervision, writing – review & editing. All authors agreed with the results and conclusions.

Funding: No funding source is reported for this study.

Ethical statement: This study was conducted in accordance with the Declaration of Helsinki, in accordance with the ethical standards of the University of Split and the rules of the profession (document; class: 042-01/24-01/00020, reg. no: 2181-204-02-07-24-00003). Approval for conducting the research was also obtained from the Ministry of Science, Education and Youth and the Education and Teacher Training Agency of the Republic of Croatia. Informed consent was obtained from all subjects involved in the study. In addition, all data were anonymized to guarantee the privacy of the participants.

AI statement: The authors report that they used ChatGPT and Microsoft Copilot (OpenAI) exclusively to improve the readability and linguistic quality of the manuscript. All content was carefully reviewed and verified by the authors to ensure its accuracy and scientific integrity.

Declaration of interest: No conflict of interest is declared by the authors.

Data sharing statement: Data supporting the findings and conclusions are available upon request from the corresponding author.

REFERENCES

Afif, N. F., Nugraha, M. G., & Samsudin, A. (2017). Developing energy and momentum conceptual survey (EMCS) with four-tier diagnostic test questions. *AIP Conference Proceedings, 1848*(1), Article 050010. https://doi.org/10.1063/1.4983966

Almisad, B. M., & Aleidan, A. A. (2025). Perceptions of gamification in education: Evidence from a developing country context. *Eurasia Journal of Mathematics, Science and Technology Education, 21*(10), Article em2718. https://doi.org/10.29333/ejmste/17177

Alonzo, S. M. D., & Mistades, V. M. (2021). Students' conceptual understanding and problem solving of the work energy and impulse momentum theorems in a flipped classroom. *Journal of Physics: Conference Series, 1882*, Article 012003. https://doi.org/10.1088/1742-6596/1882/1/012003

Brundage, M. J., Maries, A., & Singh, C. (2023). Using the energy and momentum conceptual survey to investigate progression in student understanding from introductory to advanced levels. *Physical Review Physics Education Research, 19*, Article 020132. https://doi.org/10.1103/PhysRevPhysEducRes.19.020132

Dega, B. G., & Govender, N. (2016). Assessment of students' scientific and alternative conceptions of energy and momentum using concentration analysis. *African Journal of Research in Mathematics, Science and Technology Education, 20*(3), 201-213. https://doi.org/10.1080/18117295.2016.1218657

Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use Welch's t-test instead of student's t-test. *International Review of Social Psychology, 30*(1), 92-101. https://doi.org/10.5334/irsp.82

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining "gamification". In *Proceedings of the 15th International Academic MindTrek Conference* (pp. 9-15). https://doi.org/10.1145/2181037.2181040

Dichev, C., & Dicheva, D. (2017). Gamifying education: What is known, what is believed and what remains uncertain: A critical review. *International Journal of Educational Technology in Higher Education, 14*, Article 9. https://doi.org/10.1186/s41239-017-0042-5

Dichev, C., Dicheva, D., & Ismailova, R. (2022). Motivators matter when gamifying learning activities. In M. E. Auer, W. Pachatz, & T. Rüütmann (Eds.), *Learning in the age of digital and green transition. ICL 2022. Lecture notes in networks and systems, vol 633* (pp. 947-958). Springer. https://doi.org/10.1007/978-3-031-26876-2_89

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*(1), 109-132. https://doi.org/10.1146/annurev.psych.53.100901.135153

Fernando, P. A., & Premadasa, H. S. (2024). Use of gamification and game-based learning in educating generation alpha: A systematic literature review. *Educational Technology & Society, 27*(2), 114-132. https://doi.org/10.30191/ETS.202404_27(2).RP03

Forndran, F., & Zacharias, C. R. (2019). Gamified experimental physics classes: A promising active learning methodology for higher education. *European Journal of Physics, 40*(4), Article 045702. https://doi.org/10.1088/1361-6404/ab215e

Gaurina, M., & Pavlin, J. (2025). Teachers' knowledge and beliefs about gamification in physics classrooms. *Journal of Baltic Science Education, 24*(2), 255-270. https://doi.org/10.33225/jbse/25.24.255

Hake, R. R. (1998). Interactive engagement versus traditional methods: A six thousand student survey of mechanics test data. *American Journal of Physics, 66*(1), 64-74. https://doi.org/10.1119/1.18809

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? — A literature review of empirical studies on gamification. In *Proceedings of the 47th Hawaii International Conference on System Sciences* (pp. 3025-3034). https://doi.org/10.1109/HICSS.2014.377

Ibáñez, M.-B., Di-Serio, Á., & Delgado-Kloos, C. (2014). Gamification for engaging computer science students in learning activities: A case study. *IEEE Transactions on Learning Technologies, 7*(3), 291-301. https://doi.org/10.1109/TLT.2014.2329293

Maries, A., Brundage, M., & Singh, C. (2023). *Upper-level students' conceptual understanding of energy and momentum* [Paper presentation]. The Physics Education Research Conference 2023. https://doi.org/10.1119/perc.2023.pr.Maries

Ministarstvo znanosti i obrazovanja Republike Hrvatske. (2019). Kurikulum nastavnog predmeta Fizika za osnovne škole i gimnazije u Republici Hrvatskoj [Physics curriculum for primary schools and grammar schools in the Republic of Croatia]. *Narodne Novine*. https://narodne-novine.nn.hr/clanci/sluzbeni/2019_01_10_193.html

Murillo Zamorano, L. R., López Sánchez, J. Á., Godoy Caballero, A. L., & Bueno Muñoz, C. (2021). Gamification and active learning in higher education: Is it possible to match digital society, academia and students' interests? *International Journal of Educational Technology in Higher Education, 18*, Article 13. https://doi.org/10.1186/s41239-021-00249-y

Opitz, S. T., Neumann, K., Bernholt, S., & Harms, U. (2017). How do students understand energy in biology, chemistry, and physics? Development and validation of an assessment instrument. *Eurasia Journal of Mathematics, Science and Technology Education, 13*(7), 3019-3042. https://doi.org/10.12973/eurasia.2017.00703a

Ortiz Rojas, M., Chiluiza, K., Valcke, M., & Bolanos Mendoza, C. (2025). How gamification boosts learning in STEM higher education: A mixed methods study. *International Journal of STEM Education, 12*, Article 1. https://doi.org/10.1186/s40594-024-00521-3

Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS* (7th ed.). Routledge. https://doi.org/10.4324/9781003117452

Radnai, T., Bónis, B., Pálosi, L., & Juhász, I. (2019). Educational experiments with motion simulation programs: Can gamification be effective in teaching mechanics? *Journal of Physics: Conference Series, 1223*, Article 012006. https://doi.org/10.1088/1742-6596/1223/1/012006

Richter, K., & Kickmeier-Rust, M. (2025). Gamification in physics education: Play your way to better learning. *International Journal of Serious Games, 12*(1), 59-81. https://doi.org/10.17083/ijsg.v12i1.858

Rose, J. A., O'Meara, J. M., Gerhardt, T. C., & Williams, M. (2016). Gamification: Using elements of video games to improve engagement in an undergraduate physics class. *Physics Education, 51*(5), Article 055007. https://doi.org/10.1088/0031-9120/51/5/055007

Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review, 3*(1), 16-20. https://doi.org/10.1007/BF02322211

Ruxton, G. D. (2006). The unequal variance *t*-test is an underused alternative to student's *t*-test and the Mann-Whitney *U* test. *Behavioral Ecology, 17*(4), 688-690. https://doi.org/10.1093/beheco/ark016

Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review, 32*, 77-112. https://doi.org/10.1007/s10648-019-09498-w

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*(1), 107-120. https://doi.org/10.1007/s11336-008-9101-0

Singh, C., & Rosengrant, D. (2003). Multiple-choice test of energy and momentum concepts. *American Journal of Physics, 71*(6), 607-617. https://doi.org/10.1119/1.1571832

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*(6), 1273-1296. https://doi.org/10.1007/s11165-016-9602-2

Triantafyllou, S. A., Georgiadis, C., & Sapounidis, T. (2025). Gamification in education and training: A literature review. *International Review of Education, 71*, 483-517. https://doi.org/10.1007/s11159-024-10111-8

Vrcelj, A., Hoic Božić, N., & Dlab, M. H. (2023). Use of gamification in primary and secondary education: A systematic literature review. *International Journal of Educational Methodology, 9*(1), 13-27. https://doi.org/10.12973/ijem.9.1.13

Wang, A. I., & Lieberoth, A. (2016). The effect of points and audio on concentration, engagement, enjoyment, learning, motivation, and classroom dynamics using Kahoot! In T. Connolly, & L. Boyle (Eds.), *Proceedings of the 10th European Conference on Games Based Learning* (pp. 737-748).

Wu, X., Guthrie, M. W., Scanlon, E. M., & Li, Y. (2023). Analyzing the dimensionality of the energy and momentum conceptual survey using question response theory. In *Proceedings of the Physics Education Research Conference 2023* (pp. 397-402). https://doi.org/10.1119/perc.2023.pr.Wu

Yana, A. U., Suparmi, S., Sarwanto, S., & Widha, S. (2022). Online-based assessment of students' conceptual understanding of energy and momentum using Quizizz. *Journal of Physics: Conference Series, 2392*, Article 012011. https://doi.org/10.1088/1742-6596/2392/1/012011