INTERNATIONAL MULTIDISCIPLINARY JOURNAL OF RESEARCH FOR INNOVATION, SUSTAINABILITY AND EXCELLENCE
Published by: RISEJOURNALS.ORG | Online ISSN: 3028-032X | Print ISSN: 3028-0370
VOLUME 5, ISSUE NO. 5

Problem-Based Gamified Teaching Strategy in Percentage Composition and Formula of Compounds

DOI: 10.5281/zenodo.24004910

MARIMAR E. TONILON
esparciamarimar@gmail.com
Cebu Technological University, Philippines

Abstract:

This study examined the effectiveness of a Problem-Based Gamified Teaching Strategy in teaching percentage composition and formula of compounds to Grade 11 Senior High School students at Don Sergio Osmena Sr. Memorial National High School, Bliss, Labangon, Cebu City, during the second quarter of the first semester of School Year 2025-2026. Using a quasi-experimental, pre-test-post-test control-group design combined with a descriptive survey, 44 Academic Track-Medical students participated: 20 in the control group and 24 in the experimental group. Both groups received traditional lecture-based instruction; the experimental group additionally experienced the Box Breaker Race, a non-digital, tabletop gamified problem-based activity with four levels of increasing difficulty, points, badges, teamwork, and immediate feedback. Achievement was measured with a 30-item researcher-made test aligned to three Department of Education competencies, and lifelong learning motivation was measured with an 11-item Lifelong Learning Motivation Scale adapted from Usta (2023). At baseline, both groups Did Not Meet Expectations. After instruction, the experimental group reached Outstanding ($M = 28.04$, $SD = 2.73$, 93.47%), while the control group reached Very Satisfactory ($M = 25.50$, $SD = 3.94$, 85.00%). Paired t-tests showed significant gains in both groups ($p < .001$), but the experimental group's mean gain (15.79) significantly exceeded the control group's (11.15), $t = 3.76$, $p = .001$, $d = 1.16$. Motivation was Strongly Agree overall ($M = 3.25$), highest in the cognitive domain ($M = 3.48$). Post-test scores and motivation were not significantly correlated ($r = -0.028$, $p = .898$). The strategy is therefore recommended as an effective, low-cost, non-digital intervention for conceptually and computationally demanding chemistry topics.

Keywords: problem-based learning, gamification, percentage composition, empirical and molecular formula, chemistry education, lifelong learning motivation

Introduction:

Education remains one of the most powerful drivers of social and economic progress, yet its promise is unevenly realised. Globally, poverty, teacher shortages, limited connectivity, and scarce instructional resources continue to constrain learning, and the COVID-19 pandemic deepened those inequalities. In the Philippines, the abrupt shift to blended and distance learning left many learners struggling with motivation, self-regulation, and conceptual understanding. As face-to-face classes resumed, students returned with uneven prior knowledge, fragile study routines, and varied access to support at home. Teachers, in turn, faced the difficult task of rebuilding foundational skills while also covering demanding, higher-order competencies on schedule. Within this recovery context, the quality of day-to-day classroom strategy matters more than ever, because it is the mechanism through which curriculum intentions become actual learning.

Chemistry exemplifies this challenge. It requires learners to move fluently among macroscopic phenomena, submicroscopic models, and symbolic representations, often called the chemistry triplet, while also performing precise calculations. Percentage composition and formula of compounds sits squarely at that intersection. To succeed, a student must recall atomic masses, compute molar mass, apply proportional reasoning, convert percentage composition into mole ratios, simplify to an empirical formula, and then use molar mass to derive a molecular formula. Each step is modest in isolation, but the chain is long, and a single arithmetic or conceptual slip can invalidate the result. When instruction presents this chain only as a set of worked examples to copy, many learners memorise procedures without understanding why each step is needed, and they struggle to transfer the procedure to unfamiliar, real-life problems.

In the Philippine Senior High School Chemistry 1 curriculum, learners are expected to make accurate measurements, relate electronic structure to the behaviour of particles, solve percentage composition and compound formulas, and explain chemical bonding and properties of compounds. The second quarter therefore includes atomic mass, percentage composition and formula of compounds, chemical nomenclature, and chemical bonds. The three competencies addressed in this study are to calculate the percent composition of a compound from its formula, to use a table to show similarities and differences between empirical and molecular formulas, and to calculate the empirical and molecular formula of a given compound. These are higher-order competencies because they require application, analysis, and evaluation, not recall alone. A teaching strategy for this topic must therefore do two things at once: secure procedural accuracy and build conceptual understanding that can be applied beyond the worksheet.

This study responds to that dual demand by combining two well-established approaches into one coherent strategy. Problem-based learning (PBL) organises learning around authentic, ill-structured problems that require inquiry, collaboration, and decision-making. Gamification adds game elements, such as points, levels, badges, immediate feedback, teamwork, and friendly competition, to non-game learning tasks. Separately, each approach has shown promise. Together, they offer a plausible solution to a specific classroom problem: how to make a cognitively demanding, calculation-heavy topic feel purposeful, social, and manageable without requiring laptops, internet access, or expensive materials. The resulting intervention, called the Box Breaker Race, is deliberately non-digital, because the study site is a public school where device availability cannot be assumed.

The purpose of this study was therefore to determine the effectiveness of the Problem-Based Gamified Teaching Strategy in teaching percentage composition and formula of compounds to Grade 11 students. Specifically, it compared pre-test and post-test performance of control and experimental groups across the three competencies, tested the significance of within-group and between-group gains, examined students' lifelong learning motivation in cognitive, affective, and psychomotor domains, and tested whether post-test performance was related to motivation. On the basis of the findings, it proposes an enhanced version of the strategy for wider classroom use. The study is anchored in Republic Act No. 10533, the Enhanced Basic Education Act of 2013, which promotes learner-centred, inquiry-driven, and contextualised instruction, and in the Strengthened Senior High School Curriculum (2025), which emphasises application, flexibility, and practical skills.

The significance of the study is practical as well as theoretical. For students, it offers a more engaging route into a topic that many experience as abstract and intimidating. For science teachers, it provides a ready-to-use, low-cost lesson design for topics that combine computation and concept. For school administrators, it offers evidence about whether a modest change in instructional design can produce meaningful gains without new technology. For parents, it illustrates how classroom learning can be connected to everyday decisions about food, medicine, and household products. For the educational system, it contributes evidence on whether combining PBL and gamification can advance both achievement and the dispositions associated with lifelong learning, which the curriculum explicitly seeks to cultivate.

Literature Review:

Problem-Based Learning: From Theory to Chemistry Classrooms

Problem-based learning is a student-centred pedagogy in which learning begins with an authentic problem rather than with exposition. Learners identify what they know, what they need to know, and how to find out, then test their emerging understanding by proposing and defending a solution. Azer (2001) argued that PBL aligns educational objectives with the reasoning that professionals actually use, while Bell (2012) showed that it can help learners take ownership of their learning. Ostby (2022) found that PBL supported critical thinking and peer relationships, and Alreshidi and Lally (2024) reported that teacher preparation in PBL implementation mattered for student outcomes. In science, Bagay et al. (2023) developed a PBL lesson plan for chemistry that used real-life problems to promote inquiry, independent thinking, knowledge construction, and self-directed learning. Andal and Hermosa (2024) likewise found that PBL strategies supported creative thinking among Grade 12 students. Taken together, this literature suggests that PBL is not merely a motivational device; it changes the cognitive work that students do, from reproducing procedures to constructing and justifying solutions. For percentage composition and formula of

compounds, that shift is crucial, because the topic rewards learners who can explain why a mole ratio must be simplified, why molar mass determines the molecular formula, and why a result is chemically plausible, not only those who can execute a memorised algorithm.

Gamification and Non-Digital Gamification

Gamification is the use of game-design elements in non-game contexts to influence motivation, engagement, and behaviour. Common elements include points, levels, badges, leaderboards, challenges, immediate feedback, and narrative. Chans and Portuguez Castro (2021) found that gamification increased motivation and engagement among higher-education chemistry students. Smiderle et al. (2020) showed that gamification affected learning, engagement, and behaviour, with effects moderated by personality traits, which cautions against a one-size-fits-all design. Li, Ma, and Shi (2023) synthesised experimental evidence in a meta-analysis and concluded that gamification can promote teaching and learning, although effect sizes depend on design quality and alignment with learning goals. Kulkarni et al. (2022) and Hou (2023) emphasised that the choice of game and its fit to the learning objective matter more than the mere presence of points or competition. A crucial distinction in this literature is between game-based learning, in which a game itself is the instructional medium, and gamification, in which game elements are layered onto an existing learning task. This study adopts the second approach, preserving the chemistry problem as the core task while adding structure, feedback, and incentive around it.

A second distinction is between digital and non-digital gamification. Much of the international literature assumes access to platforms, devices, and connectivity. That assumption does not hold in many Philippine public schools. Zainuddin and Keumala (2021) argued that gamification without digital platforms can motivate learners at home during the pandemic, and Hashim et al. (2023) systematically reviewed board-game approaches in science education and found that they can support interest, collaboration, and conceptual engagement. Sousa et al. (2023) reviewed board, tabletop, and other analogue game-based learning approaches and concluded that well-designed analogue games can support learning when they are tightly aligned to curriculum goals. Taspinar, Schmidt, and Schuhbauer (2016) similarly showed that a board-game approach can support knowledge acquisition. The present study builds directly on this strand by using printed boxes, whiteboards or manila paper, badges, and scoring sheets rather than software. Non-digital design is not a limitation to apologise for; it is a deliberate equity choice that makes the intervention usable where it is most needed.

Theoretical Foundations

Four theoretical perspectives frame the intervention. Self-Determination Theory (Ryan and Deci, 2000) holds that motivation is strongest when the needs for autonomy, competence, and relatedness are satisfied. In the Box Breaker Race, autonomy appears in team decisions about strategy and solution pathways, competence appears in leveled challenges that are difficult but attainable, and relatedness appears in teamwork and shared goals. Constructivist theory, associated with Bruner and Piaget, holds that learners actively construct knowledge by connecting new ideas to prior understanding. PBL operationalises this principle by requiring students to build the solution rather than receive it. Experiential Learning Theory (Kolb, 2001) adds a cycle of concrete experience, reflective observation, abstract conceptualisation, and active experimentation. In the game, students experience a problem, reflect on what worked, abstract the chemical principle, and test it again at the next level. Behaviourism (Skinner, 1988) explains the complementary role of reinforcement: points, badges, and recognition increase the likelihood that effortful, accurate work will be repeated. Keller's ARCS Model, which organises motivation around Attention, Relevance, Confidence, and Satisfaction, guided the design of scenarios that capture attention, connect chemistry to everyday life, build confidence through graduated difficulty, and provide satisfying closure through winning and recognition.

Percentage Composition and Formula of Compounds: Why Learners Struggle

Research on chemistry learning has long documented the difficulty of stoichiometry and related calculations. Johnstone (2010) argued that learners struggle because chemistry simultaneously demands macroscopic, submicroscopic, and symbolic thinking. Taber (2013) revisited the chemistry triplet and showed that novices often cannot coordinate these levels. BouJaoude and Barakat (2003) found that students' problem-solving strategies in stoichiometry were related to conceptual understanding and learning approaches, while Chandrasegaran, Treagust,

and Mocerino (2007) documented difficulties in moving among multiple representations. Nieswandt (2007) linked affect and conceptual understanding, suggesting that anxiety and low confidence can compound cognitive difficulty. Towns et al. (2012) examined how external representations shape reasoning in biochemistry, reinforcing the value of tabular and visual supports. Wu, Krajcik, and Soloway (2001) found that visualisation tools helped students use chemical representations more productively. For the present topic, these findings imply that instruction should not treat calculation and representation as separate. A table that compares empirical and molecular formulas, a worked mole-ratio chain, and a real-life scenario that makes the numbers meaningful are complementary supports, not optional extras.

Cognitive Load, Representation, and Scaffolding

Cognitive Load Theory (Sweller, 2011; Sweller, van Merrienboer, and Paas, 2019) holds that working memory is limited, so instruction should reduce extraneous load, manage intrinsic load, and promote germane load. Leveled problems serve this function: an easy problem establishes the procedure, a medium problem adds a representational demand, a hard problem requires a full empirical-to-molecular chain, and an advanced problem requires evaluation and ethical reasoning. Ainsworth (2006) provided a framework for learning with multiple representations, arguing that different representations can complement, constrain, and help learners construct deeper understanding. Mayer (2014) similarly emphasised that learners build understanding when they can integrate verbal and visual information. In this study, tables, formulas, scenarios, and team discussion are intentionally combined so that a learner who does not grasp the idea from a formula alone may grasp it from a table or from a peer's explanation. Although the original thesis sometimes used the language of differentiated instruction, the intervention tested here is more precisely described as a gamified PBL strategy with graduated difficulty and collaborative scaffolding. That clarification matters for interpretation: gains are attributed to the combined PBL-gamification design, not to a separately implemented differentiated-instruction programme.

Lifelong Learning Motivation

Lifelong learning motivation refers to a learner's sustained, self-directed desire to acquire knowledge and skills beyond formal schooling, for personal growth, career development, and civic participation. Usta (2023) developed and validated a Lifelong Learning Motivation Scale (LLMS) that captures cognitive, affective, and psychomotor dimensions. The cognitive dimension concerns the desire to acquire new knowledge and to persist through difficulty. The affective dimension concerns attitudes toward learning, openness to new topics, and confidence. The psychomotor dimension concerns the sharing and application of knowledge and skills in everyday life. In chemistry, these dimensions are not abstract. A student who checks a food label, questions a medicine dosage, or explains a household chemical to a family member is enacting lifelong learning. The present study therefore treats motivation not as a pleasant side-effect of good teaching but as an outcome in its own right, worthy of direct measurement. It also recognises that motivation and achievement need not move together in a simple linear way: a student can feel more motivated without yet scoring higher, and a student can score higher through drill without developing a durable desire to keep learning. That possibility is tested explicitly in the correlational analysis.

Synthesis and Research Gap

The literature supports each component of the intervention, but few studies test their combination, in a non-digital form, on this specific chemistry topic, in a Philippine public-school context, with both achievement and lifelong motivation as outcomes. PBL studies often focus on higher education or on general science, gamification studies often assume digital platforms, and chemistry studies often focus on achievement alone. The present study fills that gap by testing a low-cost, curriculum-aligned, non-digital Problem-Based Gamified Teaching Strategy on percentage composition and formula of compounds, with a control-group comparison and a validated motivation measure. Its conceptual framework is summarised in Figure 1. The framework predicts that real-life problems, processed through leveled gamified teamwork with feedback and reinforcement, will produce higher achievement and stronger lifelong motivation than lecture alone, and that the effect will be visible across all three competencies, with the largest advantage where conceptual and procedural demands are most tightly coupled.

INTERNATIONAL MULTIDISCIPLINARY JOURNAL OF RESEARCH FOR INNOVATION, SUSTAINABILITY AND EXCELLENCE
Published by: RISEJOURNALS.ORG | Online ISSN: 3028-032X | Print ISSN: 3028-0370
VOLUME 5, ISSUE NO. 5

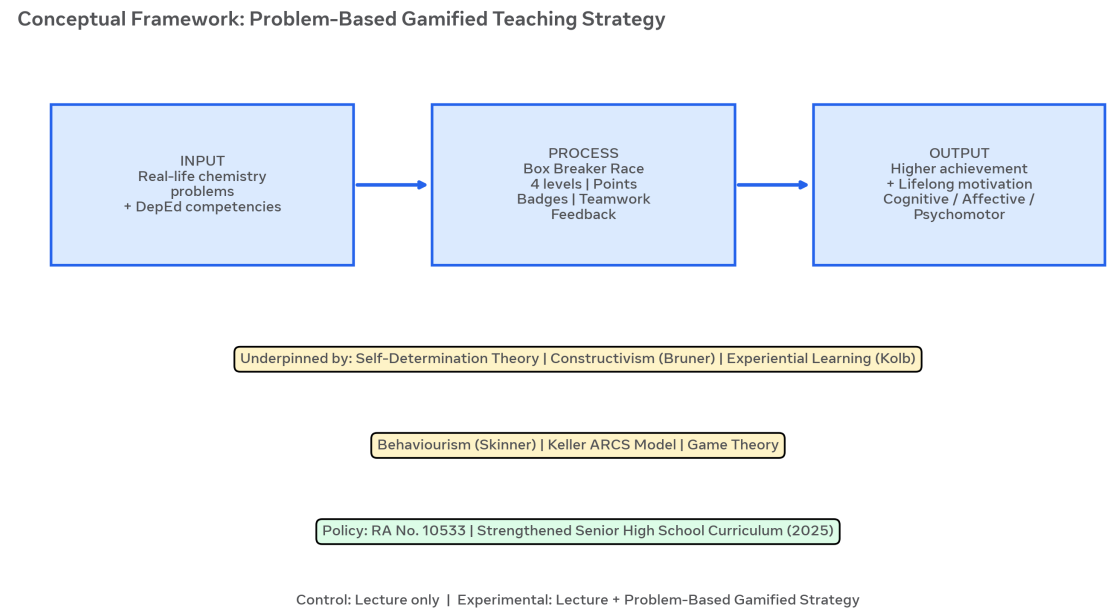


Figure 1. Conceptual framework of the study (Input-Process-Output).

Methodology:

Research Design

This study used a quasi-experimental, pre-test-post-test control-group design, combined with a descriptive survey. A quasi-experimental design was appropriate because the study was conducted in intact classes; students could not be randomly assigned individually without disrupting school organisation. The control group received traditional lecture-based instruction only. The experimental group received the same lecture plus the Problem-Based Gamified Teaching Strategy. Both groups completed the same pre-test and post-test, and both completed the Lifelong Learning Motivation Scale after instruction. The descriptive component summarised motivation levels by domain. The design allowed two complementary inferences: whether each group improved, and whether the experimental group improved more.

Research Environment and Participants

The study was conducted at Don Sergio Osmena Sr. Memorial National High School, Bliss, Labangon, Cebu City, a public school founded in 1984 that offers Junior and Senior High School programmes. Participants were 44 Grade 11 Academic Track-Medical students in two intact sections enrolled in Chemistry 1 during the second quarter of the first semester of School Year 2025-2026. The control group had 20 students; the experimental group had 24 students. The sample was purposive and complete for the two sections, which limits generalisation but strengthens ecological validity: the intervention was tested in real classes, with real time constraints, rather than in a laboratory setting.

Instruments

Achievement was measured with a 30-item multiple-choice test, with 10 items per competency: (1) calculating percent composition from a formula, (2) using a table to show similarities and differences between empirical and molecular formulas, and (3) calculating empirical and molecular formulas of a given compound. Items were adapted from Department of Education Senior High School modules (2023) and from Walsh (2020), and were aligned to a Table of Specifications. Each correct answer scored 1 point; each incorrect answer scored 0. Motivation was measured with an 11-item Lifelong Learning Motivation Scale (LLMS) adapted from Usta (2023), with five cognitive items, three affective items, and three psychomotor items, rated on a 4-point scale from 1 (Strongly Disagree) to 4 (Strongly Agree). Parental consent was obtained, and student responses were treated confidentially.

INTERNATIONAL MULTIDISCIPLINARY JOURNAL OF RESEARCH FOR INNOVATION, SUSTAINABILITY AND EXCELLENCE
Published by: RISEJOURNALS.ORG | Online ISSN: 3028-032X | Print ISSN: 3028-0370
VOLUME 5, ISSUE NO. 5

Intervention

The Box Breaker Race. The experimental group was divided into teams of five to six students. Four boxes were arranged as stations, each containing one or two problem-based scenarios of increasing difficulty: Box 1 (Easy) asked for the percent composition of aspirin, C9H8O4; Box 2 (Medium) asked for the percent composition of calcium carbonate, CaCO3, and completion of a comparison table; Box 3 (Hard) presented a forensic scenario requiring calculation of empirical and molecular formulas from percentage data and molar mass; and Box 4 (Advanced) required discussion of safe handling, toxicity, and societal measures to prevent misuse of toxic substances. Teams solved each problem collaboratively, wrote their final answer on a whiteboard or manila paper, and raised it for checking. A correct answer allowed the team to advance; an incorrect answer required retry. The first team to complete all four boxes won, with points, badges, and recognition. No time limit was imposed, to prioritise accuracy and reasoning over speed alone.

Data-Gathering Procedure

After approval from the school principal and coordination with the cooperating teacher, the researcher administered the pre-test to both groups, delivered the traditional lecture to both groups, and then administered the post-test and LLMS. The gamified intervention was then implemented with the experimental group, followed by the post-test and LLMS for that group, consistent with the flow in the original thesis. Data were tallied, checked for completeness, and subjected to statistical analysis. Ethical safeguards included voluntary participation, parental consent, anonymity in reporting, and the use of the intervention only as a supplement to, not a replacement for, the required lecture.

Data Analysis

Weighted means and standard deviations summarised pre-test, post-test, and motivation scores. Paired-samples t-tests tested within-group pre-test to post-test change. Independent-samples t-tests compared mean gains between groups. Pearson product-moment correlation tested the relationship between post-test scores and motivation. All tests used alpha = .05. Effect sizes were reported as Cohen's d, interpreted as small (0.20), medium (0.50), and large (0.80) following Cohen (1988). Achievement was interpreted using Department of Education Order No. 8, s. 2015: 90% and above Outstanding, 85-89% Very Satisfactory, 80-84% Satisfactory, 75-79% Fairly Satisfactory, and below 75% Did Not Meet Expectations. Motivation was interpreted as 3.26-4.00 Strongly Agree, 2.51-3.25 Agree, 1.76-2.50 Disagree, and 1.00-1.75 Strongly Disagree.

Table 1. Frequency distribution of respondents

Group	Male (n)	Female (n)	Total	Share
Control (Group B)	0	20	20	45.5%
Experimental (Group A)	6	18	24	54.5%
Total	6	38	44	100%

Note. Distribution as reported in the original thesis. The marked gender imbalance reflects the intact composition of the two sections and is a limitation for generalisation.

Table 2. Scoring procedure for achievement (DepEd Order No. 8, s. 2015)

Percentage	Category	Interpretation
90 and above	Outstanding	Mastered all competencies
85-89	Very Satisfactory	Mastered at least two competencies
80-84	Satisfactory	Mastered at least one competency
75-79	Fairly Satisfactory	Requires moderate supervision
Below 75	Did Not Meet Expectations	Competencies not met; needs immediate support

Table 3. Scoring procedure for Lifelong Learning Motivation Scale

Range	Category	Interpretation
3.26-4.00	Strongly Agree	Very positive toward the strategy
2.51-3.25	Agree	Positive toward the strategy
1.76-2.50	Disagree	Negative toward the strategy
1.00-1.75	Strongly Disagree	Very negative toward the strategy

Results and Discussion:

Pre-Test Performance: A Low, Comparable Baseline

Table 4 shows that, before instruction, both groups performed well below the passing threshold. The control group averaged 14.35 (SD = 4.36, 47.83%) and the experimental group averaged 12.25 (SD = 4.26, 40.83%), both in the Did Not Meet Expectations category. The 2.10-point difference is small relative to the standard deviations and to the 30-point scale, and it favours the control group. This pattern is important for internal validity: the experimental group did not start ahead. If anything, it started slightly behind, so its later superiority cannot be explained by a pre-existing advantage. In quasi-experimental research with intact classes, baseline equivalence is never perfect, but the present pattern is reassuring because the direction of the small baseline difference works against the hypothesis. A study in which the experimental group began higher and ended higher would leave open the explanation that the groups were simply different. Here, the experimental group began lower and ended higher, which strengthens the inference that the intervention, rather than prior ability, contributed to the difference.

Competency-level results clarify where the difficulty lay. In Competency 1, calculating percent composition from a formula, the control group averaged 3.95 (39.50%) and the experimental group 3.00 (30.00%). This was the weakest competency for both groups. In Competency 2, using a table to compare empirical and molecular formulas, the control group averaged 5.70 (57.00%) and the experimental group 4.29 (42.90%). This was the strongest competency for both groups, though still far below passing. In Competency 3, calculating empirical and molecular formulas, the control group averaged 4.70 (47.00%) and the experimental group 4.96 (49.60%), the only competency in which the experimental group began slightly ahead. The pattern across competencies is theoretically coherent. Competency 2 draws on comparison, classification, and table reading, skills that students encounter in many subjects and that can be performed even with incomplete computational fluency. Competency 1 and Competency 3 require a chain of calculations in which each step depends on the previous one. A student who misremembers an atomic mass, miscalculates molar mass, or mishandles a ratio will arrive at a wrong answer even if the underlying concept is partly understood. The baseline data therefore suggest that the primary barrier was not unfamiliarity with vocabulary alone, but fragile procedural fluency and weak integration of concept and calculation.

These baseline findings align with a substantial body of research on chemistry learning. Johnstone (2010) argued that chemistry is difficult because learners must coordinate macroscopic, submicroscopic, and symbolic levels simultaneously. Taber (2013) showed that novices often cannot move fluidly among these levels, so a problem that appears routine to a teacher can impose a heavy cognitive load on a student. BouJaoude and Barakat (2003) found that students' strategies in stoichiometry were tied to their conceptual understanding, and Chandrasegaran et al. (2007) documented persistent difficulties with multiple representations. Nieswandt (2007) added that affect matters: students who expect chemistry to be difficult may approach calculations with anxiety that further reduces working-memory capacity. The low pre-test scores in this study are therefore not evidence of a uniquely weak cohort; they are consistent with what the literature predicts when a demanding topic is assessed before systematic instruction. They also justify the intervention: if students had already mastered the topic, there would be little room for a teaching strategy to make a difference, and a ceiling effect would obscure any effect. Here, there was ample room for growth, and the question was whether instruction, and which kind of instruction, could produce it.

The variability in pre-test scores also deserves attention. Standard deviations of 4.36 and 4.26 on a 30-point scale indicate substantial heterogeneity within each class. Some students were close to passing, while others scored very low. In a traditional lecture, a teacher typically pitches instruction to the middle of the distribution, which can leave struggling learners behind and bore advanced learners. The gamified PBL design, with leveled problems and teamwork, was intended to address this heterogeneity by allowing teams to distribute cognitive work, discuss alternative approaches, and receive feedback at each level before advancing. Whether that design actually reduced the

INTERNATIONAL MULTIDISCIPLINARY JOURNAL OF RESEARCH FOR INNOVATION, SUSTAINABILITY AND EXCELLENCE
Published by: RISEJOURNALS.ORG | Online ISSN: 3028-032X | Print ISSN: 3028-0370
VOLUME 5, ISSUE NO. 5

impact of heterogeneity is examined in the post-test variability: if the experimental group's standard deviation shrinks more than the control group's, it suggests that the intervention helped bring lower-performing students up, not only pushed high performers further. That pattern is indeed observed in the post-test data, as discussed below, and it is one of the more educationally important findings of the study.

Table 4. Pre-test scores by competency

Competency (max 10)	Control M (SD)	Control % / Category	Experimental M (SD)	Experimental % / Category
1. Percent composition	3.95 (1.67)	39.50% - Did Not Meet	3.00 (1.98)	30.00% - Did Not Meet
2. Empirical vs molecular (table)	5.70 (1.75)	57.00% - Did Not Meet	4.29 (1.92)	42.90% - Did Not Meet
3. Empirical and molecular calculation	4.70 (2.45)	47.00% - Did Not Meet	4.96 (1.57)	49.60% - Did Not Meet
Overall (max 30)	14.35 (4.36)	47.83% - Did Not Meet	12.25 (4.26)	40.83% - Did Not Meet

Post-Test Performance: Both Groups Improve, Experimental Group Reaches Outstanding

Table 5 and Figure 2 show a striking change after instruction. The control group rose to 25.50 (SD = 3.94, 85.00%), classified as Very Satisfactory, while the experimental group rose to 28.04 (SD = 2.73, 93.47%), classified as Outstanding. Both groups therefore improved substantially, which is an important finding in its own right: traditional lecture-based instruction was not ineffective. It moved students from failure to a solid passing level. The experimental group, however, improved more, and its final level represents a qualitatively different outcome: mastery of all competencies, not merely mastery of most. The 2.54-point gap on the post-test is larger than the 2.10-point baseline gap in the opposite direction, so the net swing in favour of the experimental group is 4.64 points, which is the mean-gain difference analysed formally in the next section.

Competency-level post-test results show that the experimental advantage was consistent, not driven by a single competency. In Competency 1, the experimental group scored 9.38 (93.80%) compared with the control group's 8.50 (85.00%), a gap of 0.88 points. In Competency 2, the experimental group scored 9.38 (93.80%) compared with 8.55 (85.50%), a gap of 0.83 points. In Competency 3, the experimental group scored 9.29 (92.90%) compared with 8.45 (84.50%), a gap of 0.84 points. The near-uniform gap across competencies is noteworthy. It suggests that the intervention did not merely teach a trick for one type of item, but supported the underlying chain of reasoning that all three competencies require. A student who can calculate percent composition accurately is better placed to calculate an empirical formula, and a student who can compare formulas in a table is better placed to explain why a molecular formula is a multiple of an empirical formula. The consistency of the advantage supports the interpretation that the gamified PBL design strengthened integrated understanding, not isolated test-taking skill.

The reduction in variability is also informative. The experimental group's overall standard deviation fell from 4.26 at pre-test to 2.73 at post-test, while the control group's fell from 4.36 to 3.94. Both groups became more homogeneous, as expected when instruction brings lower performers up, but the experimental group became markedly more homogeneous. At the competency level, the experimental group's standard deviations on the post-test were 0.92, 1.01, and 1.20, compared with the control group's 1.40, 1.50, and 1.76. This pattern suggests that the intervention was particularly helpful for students who would otherwise have remained behind. Teamwork may be the mechanism: in a team, a student who is unsure about a step can ask, observe a peer's reasoning, and correct a misconception before the team advances. Immediate feedback at each box prevents a small error from compounding across levels. Leveled difficulty ensures that a team does not face the hardest problem before it has succeeded on easier ones, which builds confidence and procedural fluency in a graduated way. In a lecture, by contrast, a student who misses a step may remain lost for the rest of the lesson, because the lesson moves on regardless.

These post-test findings are consistent with research on active learning, representation, and cognitive load. Marzano, Pickering, and Pollock (2001) found that scaffolding and structured practice support achievement, and Plass, Moreno,

INTERNATIONAL MULTIDISCIPLINARY JOURNAL OF RESEARCH FOR INNOVATION, SUSTAINABILITY AND EXCELLENCE
Published by: RISEJOURNALS.ORG | Online ISSN: 3028-032X | Print ISSN: 3028-0370
VOLUME 5, ISSUE NO. 5

and Brunken (2010) showed that managing cognitive load improves learning in complex domains. Ainsworth (2006) argued that multiple representations, when well coordinated, help learners build robust understanding, which is directly relevant to Competency 2, where tables support comparison. Sweller (2011) argued that extraneous load should be minimised so that working memory can be devoted to germane processing; the Box Breaker Race reduces extraneous load by breaking a long procedure into four manageable boxes, each with a clear goal, and by providing feedback before the next demand is imposed. Rosenshine (2012) showed that direct, explicit instruction also produces meaningful gains, which explains why the control group improved substantially. The present study does not suggest that lecture is worthless; it suggests that lecture plus a well-designed gamified PBL supplement produces a higher level of mastery, with less variability, than lecture alone.

An alternative explanation must be considered: the experimental group may have simply spent more time on task, because the game added instructional time beyond the lecture. Time on task is a genuine confound in many intervention studies, and it cannot be fully ruled out here. However, time alone does not explain the pattern of results. Additional time spent on drill or on copying worked examples would be expected to produce gains, but not necessarily the uniform, high-level mastery observed across all three competencies, nor the marked reduction in variability. The quality of the additional time matters: it was spent solving novel, real-life problems, discussing reasoning, and receiving feedback, not merely repeating the same procedure. Future studies should record instructional time precisely and, if possible, provide the control group with an equal-time, non-gamified active-learning activity, so that the specific contribution of gamification can be isolated from the contribution of additional active practice. That design refinement is discussed in the limitations section, but it does not undermine the practical conclusion that the combined strategy, as implemented in a real school timetable, produced a meaningful advantage.

Table 5. Post-test scores by competency

Competency (max 10)	Control M (SD)	Control % / Category	Experimental M (SD)	Experimental % / Category
1. Percent composition	8.50 (1.40)	85.00% - Very Satisfactory	9.38 (0.92)	93.80% - Outstanding
2. Empirical vs molecular (table)	8.55 (1.50)	85.50% - Very Satisfactory	9.38 (1.01)	93.80% - Outstanding
3. Empirical and molecular calculation	8.45 (1.76)	84.50% - Very Satisfactory	9.29 (1.20)	92.90% - Outstanding
Overall (max 30)	25.50 (3.94)	85.00% - Very Satisfactory	28.04 (2.73)	93.47% - Outstanding

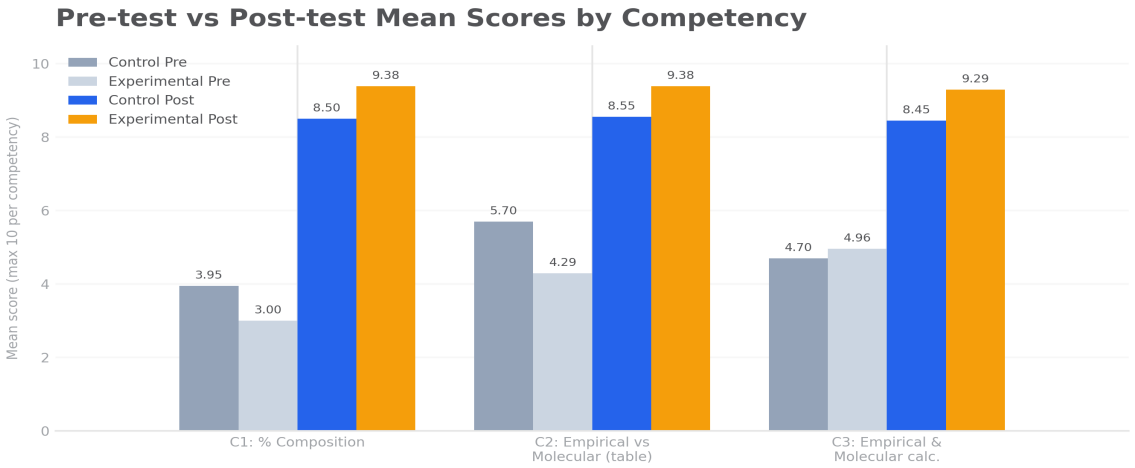


Figure 2. Pre-test versus post-test mean scores by competency and group.

Within-Group Change: Significant Gains in Both Groups, With Very Large Effects

Table 6 reports paired-samples t-tests comparing pre-test and post-test scores within each group. All comparisons were statistically significant at $p < .001$. For the control group, the overall mean difference was 11.15 points, $t = 16.00$, $d = 3.58$. For the experimental group, the overall mean difference was 15.79 points, $t = 16.40$, $d = 3.35$. According to Cohen's (1988) benchmarks, $d = 0.20$ is small, 0.50 is medium, and 0.80 is large; both groups' effects are therefore extremely large, far beyond the conventional large threshold. Such large effects are uncommon in educational research, where meta-analyses often report average effects around $d = 0.40$ to 0.80 for well-implemented interventions (Deunk et al., 2018). Several factors may explain the magnitude. First, the pre-test was administered before any instruction on the topic, so students had substantial room to grow. Second, the test was closely aligned to the taught competencies, so gains reflect taught content rather than general ability. Third, the paired design controls for individual differences, which increases statistical power. Fourth, the sample was small, and small samples can produce inflated effect-size estimates. Readers should therefore interpret the absolute magnitude with caution, while noting that the direction and significance of the effects are clear and that the between-group comparison, which is the primary test of the intervention, is more informative than the within-group magnitudes alone.

At the competency level, the experimental group's largest effect was in Competency 1 ($d = 3.19$, mean difference 6.38), followed by Competency 2 ($d = 2.25$, mean difference 5.08) and Competency 3 ($d = 2.25$, mean difference 4.33). The control group's largest effect was also in Competency 1 ($d = 3.03$, mean difference 4.55), followed by Competency 3 ($d = 1.56$, mean difference 3.75) and Competency 2 ($d = 1.40$, mean difference 2.85). The pattern is instructive. Competency 1, which was weakest at baseline, showed the largest gain in both groups, which is consistent with a ceiling-and-floor dynamic: the competency with the most room to grow often shows the largest absolute gain. However, the experimental advantage was largest in Competency 2, where the experimental effect ($d = 2.25$) exceeded the control effect ($d = 1.40$) by 0.85, compared with a 0.16 advantage in Competency 1 and a 0.69 advantage in Competency 3. This suggests that the gamified PBL design was particularly helpful for comparative, representational reasoning, the kind of reasoning that benefits most from discussion, tables, and peer explanation. In a lecture, a table comparing empirical and molecular formulas can be presented quickly, but students may not actively process the similarities and differences. In the game, teams must complete the table to advance, so they must engage with the comparison directly.

The within-group results also clarify a common misunderstanding about significance testing. A significant p-value does not, by itself, mean that an effect is educationally important; with a large enough sample, even a trivial difference can be significant. Conversely, a non-significant result does not mean that there was no effect. In this study, both groups' p-values are $< .001$, and the effect sizes are very large, so the results satisfy both statistical and practical significance criteria. The practical interpretation is straightforward: after instruction, students in both groups could do substantially more chemistry than they could before. The control group's gain shows that the standard curriculum and lecture approach, delivered by a qualified teacher, works. The experimental group's larger gain shows that adding the gamified PBL component works better. This is the kind of incremental, cumulative evidence that teachers and administrators need: not proof that one approach is perfect and the other is worthless, but evidence that a feasible modification produces a meaningful improvement over current practice.

From a theoretical perspective, the within-group gains are consistent with constructivist and experiential accounts of learning. Bruner's constructivism predicts that learners who actively build solutions will retain and transfer them better than learners who passively receive them. Kolb's experiential learning cycle predicts that concrete experience, reflection, abstraction, and experimentation, when sequenced, produce deeper learning than experience alone. The Box Breaker Race embodies that cycle: each box is a concrete experience, the team discussion is reflection, the derivation of the formula is abstraction, and the next box is experimentation with the new understanding. Behaviourist reinforcement, in the form of points and badges, may have helped sustain effort through the cycle, particularly when a team failed a box and had to retry. Self-Determination Theory predicts that this combination of autonomy, competence, and relatedness will support intrinsic motivation, which in turn supports persistence and deeper processing. The within-group gains alone cannot prove these mechanisms, but the pattern of results is consistent with them, and the motivation data provide an additional, independent test of the motivational mechanism.

Table 6. Paired-samples t-tests: pre-test versus post-test within each group

INTERNATIONAL MULTIDISCIPLINARY JOURNAL OF RESEARCH FOR INNOVATION, SUSTAINABILITY AND EXCELLENCE
Published by: RISEJOURNALS.ORG | Online ISSN: 3028-032X | Print ISSN: 3028-0370
VOLUME 5, ISSUE NO. 5

Competency	Control: Mean Diff., t, d, p	Experimental: Mean Diff., t, d, p
1. Percent composition	4.55, 13.53, 3.03, < .001*	6.38, 15.65, 3.19, < .001*
2. Empirical vs molecular (table)	2.85, 6.27, 1.40, < .001*	5.08, 11.00, 2.25, < .001*
3. Empirical and molecular calc.	3.75, 6.98, 1.56, < .001*	4.33, 11.02, 2.25, < .001*
Overall	11.15, 16.00, 3.58, < .001*	15.79, 16.40, 3.35, < .001*

*Note. *Significant at p < .05. Effect sizes are Cohen's d as reported in the original thesis.*

Between-Group Mean Gain: The Experimental Advantage Is Significant and Large

Table 7 and Figure 3 present the central test of the study: independent-samples t-tests comparing mean gains between groups. Overall, the experimental group gained 15.79 points, compared with 11.15 points for the control group, a difference of 4.64 points, t = 3.76, p = .001, d = 1.16. This is a large effect by Cohen's standards, and it translates to approximately 41.6% additional learning beyond the control group's gain. In practical terms, if a teacher taught two classes and used the gamified PBL strategy in one, that class would, on average, gain almost five more points on a 30-point test than the other class, enough to move the class average from Very Satisfactory to Outstanding. That is not a marginal improvement; it is a change in the category of mastery that the class achieves. For a school deciding whether to invest teacher time in developing gamified materials, this magnitude matters, because it suggests that the investment is likely to pay off in measurable learning, not merely in student enjoyment.

Competency-level between-group comparisons show that the advantage was significant for Competency 1 and Competency 2, but not for Competency 3. In Competency 1, the experimental gain (6.38) exceeded the control gain (4.55) by 1.83 points, t = 3.37, p = .002, d = 1.03. In Competency 2, the experimental gain (5.08) exceeded the control gain (2.85) by 2.23 points, t = 3.41, p = .001, d = 1.04. In Competency 3, the experimental gain (4.33) exceeded the control gain (3.75) by 0.58 points, t = 0.89, p = .376, d = 0.27, which is not significant and represents a small effect. This pattern requires careful interpretation. It does not mean that the intervention failed for Competency 3; both groups improved substantially on that competency, and the experimental group still scored higher on the post-test. It means that the difference in gains was not large enough, given the sample size and variability, to rule out chance. One plausible explanation is that Competency 3 is a highly structured, algorithmic task: once the procedure for converting percentage composition to empirical formula and then to molecular formula is understood, it can be executed reliably through direct instruction and practice. Kirschner, Sweller, and Clark (2006) argued that direct instruction is particularly effective for well-defined procedural tasks, and the present results are consistent with that argument. Competency 1 and Competency 2, by contrast, may benefit more from the conceptual discussion, representation, and real-life anchoring that PBL and gamification provide.

Another explanation for the non-significant Competency 3 difference is statistical power. With 20 and 24 students, the study had limited power to detect small-to-medium effects. A difference of 0.58 points with standard deviations around 1.5 to 2.0 corresponds to a small effect (d = 0.27), which would require a much larger sample to detect reliably. The study was adequately powered for the large overall effect, but not for a small competency-specific effect. Future research with a larger sample could test whether the Competency 3 advantage is real but small, or truly zero. In either case, the practical implication is the same: the intervention is most clearly beneficial for percent composition and for comparative table reasoning, and it is at least not harmful for empirical and molecular calculation. Teachers should therefore use the gamified PBL strategy as a complement to, not a replacement for, explicit procedural practice on Competency 3, ensuring that students receive both the conceptual discussion that the game provides and the structured drill that algorithmic fluency requires.

The between-group results also address a potential criticism: that the experimental group's higher post-test scores merely reflect regression to the mean or the fact that lower-scoring students have more room to grow. Regression to the mean would predict that the lower-scoring group at pre-test (the experimental group) would gain more simply because extreme scores tend to move toward the average on retesting. However, regression to the mean cannot fully explain the pattern, because the experimental group did not merely catch up to the control group; it overtook it and reached a higher category, with lower variability. Moreover, the gain difference was largest in Competency 2, where the experimental group's baseline disadvantage was largest, but also significant in Competency 1, where the baseline disadvantage was also present. A more rigorous test would use analysis of covariance (ANCOVA) with pre-test score as a covariate, which adjusts for baseline differences and provides a cleaner estimate of the treatment effect. Future

INTERNATIONAL MULTIDISCIPLINARY JOURNAL OF RESEARCH FOR INNOVATION, SUSTAINABILITY AND EXCELLENCE
Published by: RISEJOURNALS.ORG | Online ISSN: 3028-032X | Print ISSN: 3028-0370
VOLUME 5, ISSUE NO. 5

re-analysis should apply ANCOVA, and future studies should consider randomised assignment where feasible. For the present study, the combination of baseline disadvantage, post-test superiority, and large effect size makes a purely artefactual explanation unlikely, though it cannot be definitively excluded.

Table 7. Independent-samples t-tests: mean gain difference between groups

Competency	Control Gain	Experimental Gain	Difference	t	d	p
1. Percent composition	4.55	6.38	1.83	3.37	1.03	.002*
2. Empirical vs molecular (table)	2.85	5.08	2.23	3.41	1.04	.001*
3. Empirical and molecular calc.	3.75	4.33	0.58	0.89	0.27	.376
Overall	11.15	15.79	4.64	3.76	1.16	.001*

*Note. *Significant at $p < .05$.*

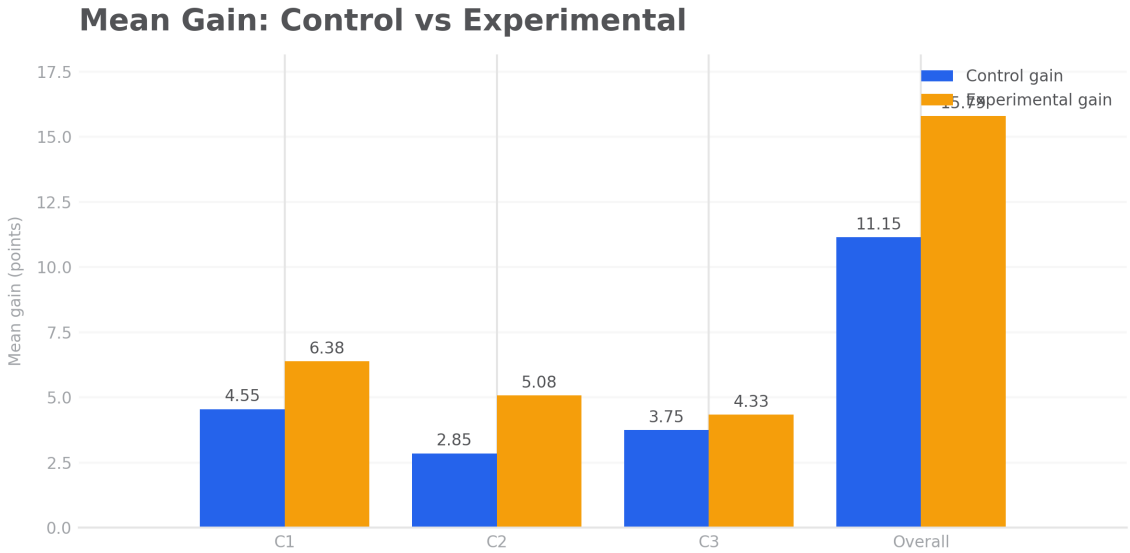


Figure 3. Mean gain by competency: control versus experimental group.

Lifelong Learning Motivation: Strongly Positive, With a Nuanced Affective Pattern

Table 8 and Figure 4 summarise students' lifelong learning motivation after instruction. The overall mean was 3.25 (SD = 0.60), at the threshold for Strongly Agree. By domain, cognitive motivation was highest at 3.48 (SD = 0.55, Strongly Agree), psychomotor motivation was 3.32 (SD = 0.56, Strongly Agree), and affective motivation was 2.96 (SD = 0.69, Agree). This ordering is meaningful. Cognitive items asked whether the strategy motivated students to acquire new knowledge, to persist through difficulty, to pursue passionate interests, to learn something new every day, and to focus on knowledge that helps achieve goals. The highest-rated item was 'I focus on acquiring new knowledge that can achieve my goals' (M = 3.88), followed by 'The difficulties I face encourage me to make more effort' (M = 3.46) and 'This activity motivates me to learn new topics' (M = 3.46). These items suggest that students did not merely enjoy the game; they experienced it as purposeful and as supporting their goals. The item about difficulties is particularly important: it suggests that challenge was experienced as motivating rather than discouraging, consistent with flow theory (Csikszentmihalyi, 1990), which holds that optimal experience occurs when challenge is matched to skill, and with Self-Determination Theory, which holds that competence is satisfying when it is earned through effort.

INTERNATIONAL MULTIDISCIPLINARY JOURNAL OF RESEARCH FOR INNOVATION, SUSTAINABILITY AND EXCELLENCE
Published by: RISEJOURNALS.ORG | Online ISSN: 3028-032X | Print ISSN: 3028-0370
VOLUME 5, ISSUE NO. 5

Psychomotor items asked whether the strategy promoted sharing knowledge with others, continued learning at every stage of life, and striving to learn new things and pursue different interests. Means were 3.38, 3.29, and 3.29 respectively, all in the Strongly Agree range. These items capture the transfer and sharing dimension of lifelong learning: not only 'I want to learn more,' but 'I want to share what I learn and keep learning throughout life.' In a team game, sharing is not an abstract ideal; it is a practical necessity. A team cannot advance unless members share knowledge, explain reasoning, and listen to alternative approaches. The psychomotor results therefore suggest that the game's collaborative structure may have translated into a disposition to share and apply knowledge beyond the classroom. Whether that disposition persists over time is an empirical question that a single post-intervention survey cannot answer, but the present results provide a promising baseline for longitudinal follow-up.

Affective motivation was lower, at 2.96 (Agree), and the pattern within the domain is revealing. Two items were Strongly Agree: 'This activity motivates me to learn new topics' (3.46) and 'This teaching strategy selects opportunities that can continuously improve myself' (3.38). One item was markedly lower: 'This method affects me negatively, even if I am not a fast learner' (M = 2.04). The wording of this item is ambiguous, and its interpretation requires care. If students interpreted it as 'The method harms me when I am slow,' then a mean of 2.04 (Disagree) means that most students disagreed that the method harmed them, which is a positive finding. If they interpreted it as a negatively worded item that should be reverse-scored, then the reported mean may not reflect the intended construct, and the affective domain mean may be underestimated. The original thesis reported the item without reverse-scoring, and the present paper preserves that reporting for fidelity, but flags the issue for future use of the scale: negatively worded items should be reverse-scored before domain means are computed, and the item's wording should be clarified in pilot testing. Even with this caveat, the affective pattern suggests that most students felt positively about the strategy, while a minority may have felt pressure from competition or from being in a team where faster peers dominated. Teachers using competitive gamification should therefore monitor team dynamics, ensure that every member has a role, and consider cooperative scoring that rewards team improvement, not only team speed, to protect slower learners' confidence.

The motivation results align with research on gamification and motivation, but they also add nuance. Chans and Portuguez Castro (2021) found that gamification increased motivation and engagement in chemistry, and Li et al. (2023) found positive effects in a meta-analysis. Smiderle et al. (2020) cautioned that effects vary by personality, which is consistent with the present finding that affective motivation was lower than cognitive motivation: not every student experiences competition in the same way. Hashim et al. (2023) and Sousa et al. (2023) found that board and tabletop games can support engagement without digital platforms, which is directly relevant to the present non-digital design. The present study extends this literature by measuring lifelong learning motivation as a multi-dimensional construct, rather than as a single enjoyment or engagement rating, and by showing that the strongest effect was on the cognitive dimension, the dimension most closely tied to sustained, self-directed learning. This is important because a strategy that is merely fun but does not build a desire to keep learning would be of limited educational value. Here, the fun appears to have been coupled with a sense of purpose and growth.

Table 8. Lifelong learning motivation (LLMS) by item and domain

Domain / Item	M	SD	Category
Cognitive: Motivates me to acquire new knowledge	3.38	0.49	Strongly Agree
Cognitive: Difficulties encourage more effort	3.46	0.51	Strongly Agree
Cognitive: Inspires effort in passionate learning	3.38	0.58	Strongly Agree
Cognitive: Gives new learning every day	3.33	0.56	Strongly Agree
Cognitive: Focus on knowledge to achieve goals	3.88	0.61	Strongly Agree

INTERNATIONAL MULTIDISCIPLINARY JOURNAL OF RESEARCH FOR INNOVATION, SUSTAINABILITY AND EXCELLENCE
Published by: RISEJOURNALS.ORG | Online ISSN: 3028-032X | Print ISSN: 3028-0370
VOLUME 5, ISSUE NO. 5

Cognitive Average	3.48	0.55	Strongly Agree
Affective: Motivates me to learn new topics	3.46	0.59	Strongly Agree
Affective: Selects opportunities to improve myself	3.38	0.49	Strongly Agree
Affective: Affects me negatively if not a fast learner*	2.04	1.00	Disagree
Affective Average	2.96	0.69	Agree
Psychomotor: Promotes sharing knowledge	3.38	0.58	Strongly Agree
Psychomotor: Promotes continued learning	3.29	0.55	Strongly Agree
Psychomotor: Strive to learn new things	3.29	0.55	Strongly Agree
Psychomotor Average	3.32	0.56	Strongly Agree
Overall	3.25	0.60	Strongly Agree

*Note. *Item wording as in the original instrument; see discussion of reverse-scoring. Scale: 3.26-4.00 Strongly Agree; 2.51-3.25 Agree; 1.76-2.50 Disagree; 1.00-1.75 Strongly Disagree.*

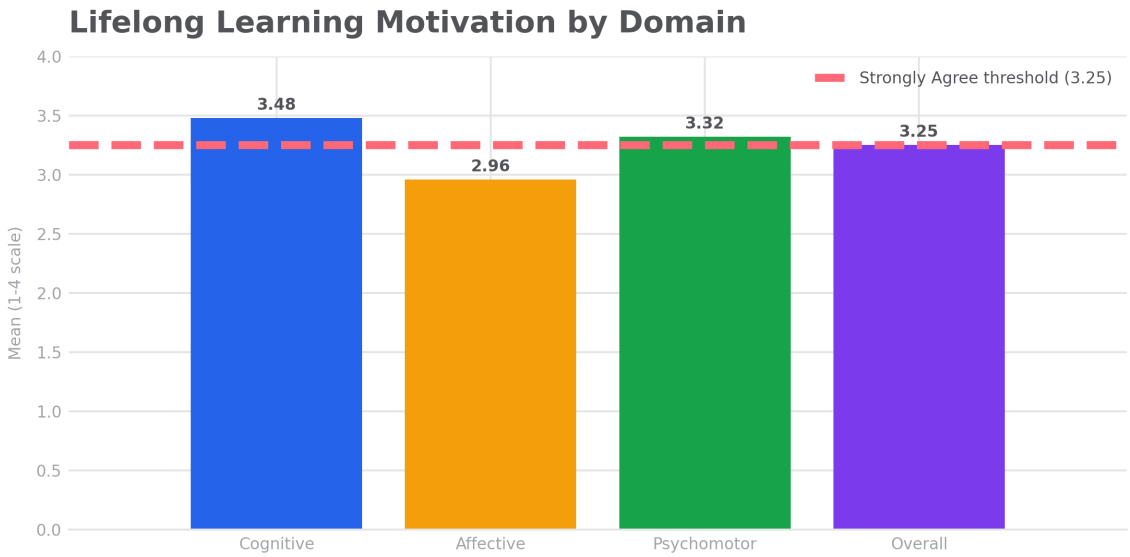


Figure 4. Lifelong learning motivation by domain, with Strongly Agree threshold.

Relationship Between Post-Test Scores and Motivation: No Significant Correlation

Table 9 shows that the correlation between post-test scores and lifelong learning motivation was r = -0.028, p = .898, which is not significant and represents essentially no relationship. The null hypothesis of no relationship is therefore retained. This finding is counterintuitive if one assumes that more motivated students should score higher, and it requires careful interpretation. First, the result does not mean that motivation is unimportant; it means that, in this sample, at this time, with these measures, individual differences in motivation did not predict individual differences in post-test scores. There are several plausible explanations, each with different implications. One is a ceiling effect: post-test scores were very high (overall mean 28.00 in the correlational analysis, with a standard deviation of 2.73), so there was limited variance for motivation to predict. When most students score near the top, the correlation is attenuated because there is little variation left to explain. A second explanation is a restriction-of-range effect in

motivation: motivation scores were also high and relatively homogeneous (M = 3.30, SD = 0.17 in the correlational table), which similarly attenuates correlation. With both variables compressed toward the top, a correlation near zero is not surprising, even if motivation and achievement are related in the broader population.

A third explanation is temporal: motivation was measured after instruction, at the same time as the post-test, so it captures a post-intervention state rather than a stable trait that could have caused achievement. Students who had just succeeded in a game may have reported higher motivation because they had succeeded, not because motivation caused their success. Conversely, students who had just struggled may have reported lower motivation for the same reason. In that case, motivation and achievement are both outcomes of the intervention, rather than one causing the other, and a cross-sectional correlation cannot disentangle them. A longitudinal design, with motivation measured before, during, and after instruction, and with follow-up achievement measured weeks later, would be needed to test causal pathways. A fourth explanation is that the relationship is non-linear or moderated by other variables. For example, motivation may matter most for students with low prior knowledge, who need persistence to catch up, but matter less for students with high prior knowledge, who can succeed through ability alone. Or motivation may predict effort and participation, which in turn predict learning, but only when combined with effective strategy use. A simple bivariate correlation cannot capture such moderated or mediated pathways. Future research should test mediation models in which the intervention affects motivation, motivation affects engagement, and engagement affects achievement, using larger samples and structural equation modelling.

The non-significant correlation also has a practical implication that teachers should not overlook. It would be a mistake to conclude that, because motivation and achievement were not correlated, teachers need not attend to motivation. The intervention improved both outcomes at the group level: the experimental group achieved more and reported strong motivation. The lack of correlation at the individual level means that, within the group, the students who scored highest were not necessarily the students who reported the highest motivation. That pattern is compatible with the idea that the game helped different students in different ways: some students may have benefited primarily through cognitive scaffolding, others through social support, and others through motivational arousal. A strategy that works through multiple pathways is more robust than one that works through a single pathway, because it can help a heterogeneous class even if no single mechanism explains all individual differences. Teachers should therefore continue to design for both achievement and motivation, while recognising that the two outcomes are not interchangeable and should be measured separately.

Methodologically, the correlational analysis should be treated as exploratory rather than definitive. The sample size of 44 provides limited power to detect small correlations; a correlation of r = 0.30, which is often considered moderate in educational research, would require a larger sample to detect reliably. The reported correlation is based on the combined sample, not on separate correlations within each group, and the original thesis does not report group-specific correlations, which could differ. For example, motivation might predict achievement in the control group, where lecture alone leaves more room for individual motivation to matter, but not in the experimental group, where the game provides external scaffolding that compensates for low motivation. Future research should report correlations separately by group, with confidence intervals, and should consider Spearman's rho if the motivation distribution is skewed. It should also check for outliers and for the influence of the reverse-worded affective item, which may have added measurement error. Despite these limitations, the present finding is valuable because it prevents an oversimplified conclusion: the strategy did not work simply because it made students more motivated, and more motivated students did not automatically score higher. Its effect is likely multi-faceted, involving cognition, representation, feedback, and social interaction as well as motivation.

Table 9. Correlation between post-test scores and lifelong learning motivation

Variable	M	SD	r	p	Decision
Post-test score	28.00	2.73	-0.028	.898	Do not reject H0
Lifelong motivation	3.30	0.17	-0.028	.898	Do not reject H0

Note. Pearson r; significant at p < .05. The post-test mean in this table (28.00) is as reported in the original thesis for the correlational analysis and differs slightly from the weighted experimental-group mean (28.04) due to rounding and sample definition.

The Enhanced Strategy: What Should Be Kept, Changed, and Added

On the basis of the findings, the enhanced Problem-Based Gamified Teaching Strategy should retain the core design that produced the gains, while addressing the weaknesses identified in the motivation and Competency 3 results. Four elements should be retained. First, the four-level structure, from easy to advanced, should be kept, because it manages cognitive load and builds confidence through graduated success. Second, the real-life scenarios, aspirin, calcium carbonate, forensic analysis, and toxic-substance safety, should be kept, because they make the chemistry purposeful and support transfer. Third, the team structure with collaborative problem-solving and immediate feedback should be kept, because it provides the social and corrective mechanisms that likely explain the reduction in variability. Fourth, the non-digital format should be kept, because it ensures equity and feasibility in public schools without reliable devices or connectivity. These are not incidental features; they are the active ingredients of the intervention, and removing any one of them would change what is being tested.

Three changes are recommended. First, for Competency 3, add a brief, explicit mini-lesson on the empirical-to-molecular algorithm before the hard box, with a worked example that students can refer to during the game. This preserves the PBL structure while ensuring that students who need direct procedural guidance receive it. The mini-lesson should be short, no more than five minutes, and should focus on the decision points where students most often err: converting percentages to grams, converting grams to moles, dividing by the smallest mole value, and multiplying by the whole-number ratio. Second, for affective motivation, add cooperative scoring that rewards improvement and helping behaviours, not only speed. For example, teams could earn bonus points for explaining their reasoning, for correcting a peer's misconception, or for improving on a previous attempt. This reduces the risk that slower learners feel penalised by competition and aligns the game's incentives with the collaborative learning that the theory predicts. Third, clarify and reverse-score the negatively worded affective item, and add two or three items that capture anxiety and confidence directly, so that future evaluations can detect whether the game is supporting or undermining slower learners' confidence.

Two additions are recommended for sustainability. First, a teacher guide should be developed that includes the curriculum guide, problem-based questionnaires, gamification materials (boards, badges, scoring sheets), a calculator and materials list, answer keys with full solutions, and common misconceptions for each box. The guide should also include a debriefing script: after the game, the teacher should lead a whole-class discussion that makes the chemical principles explicit, connects the four boxes, and addresses any errors that teams made. Without debriefing, students may remember the game but not the chemistry. Second, a follow-up assessment should be administered two to four weeks after the intervention to test retention. The present study measured immediate post-test performance, which is important, but it cannot show whether the gains persist. If the gamified PBL strategy produces durable learning, follow-up scores should remain high; if the gains fade quickly, the strategy may need spaced practice or review sessions. Future research should include this delayed post-test as a standard component, and should also collect qualitative data, such as student interviews and classroom observations, to understand how the game is experienced by different learners.

Limitations

This study has several limitations that should be considered when interpreting the findings and planning future research. First, the sample was small (N = 44) and drawn from two intact classes in a single public school, which limits statistical power and generalisability. The gender distribution was also unbalanced (38 female, 6 male), reflecting the composition of the Academic Track-Medical sections, but limiting the ability to test for gender differences. Second, assignment was quasi-experimental, not randomised, so baseline differences, teacher effects, and class culture cannot be fully excluded as explanations. The experimental group's baseline disadvantage works against the hypothesis, which is reassuring, but it does not replace randomisation. Third, the study did not control for instructional time or for the novelty effect: the game was new and exciting, and some of its effect may reflect novelty rather than the specific PBL-gamification mechanisms. A longer intervention, with multiple game sessions, would help distinguish a durable effect from a novelty effect. Fourth, the motivation scale was administered only after instruction, so pre-intervention motivation is unknown, and the correlational analysis is cross-sectional. Fifth, the original thesis contained inconsistencies in terminology, sometimes referring to differentiated instruction rather than to the gamified PBL strategy, and the present paper has standardised the terminology to the intervention actually

described and tested. Finally, the effect sizes are very large and should be interpreted with caution, given the small sample, the close alignment between test and instruction, and the possibility of inflated estimates. Replication with a larger, randomised sample, with ANCOVA, with a delayed post-test, and with equal-time control activities is needed before the strategy is recommended as a general policy.

Conclusion:

This study tested whether a Problem-Based Gamified Teaching Strategy could improve Grade 11 students' learning of percentage composition and formula of compounds, a topic that combines conceptual understanding with precise calculation. The answer, within the limits of a small quasi-experimental study in one public school, is yes. Both groups improved significantly from pre-test to post-test, which confirms that systematic instruction on this topic works. The experimental group, which received the same lecture plus the Box Breaker Race, improved significantly more, reaching Outstanding (93.47%) compared with the control group's Very Satisfactory (85.00%), with a large between-group effect ($d = 1.16$). The advantage was significant for calculating percent composition and for comparing empirical and molecular formulas in a table, and it was positive but non-significant for calculating empirical and molecular formulas, where direct procedural instruction may be equally effective. Motivation was Strongly Agree overall, highest in the cognitive domain, which suggests that students experienced the strategy as purposeful and growth-promoting, not merely entertaining.

The findings support a coherent theoretical account. When real-life problems are organised into leveled, team-based challenges with immediate feedback, points, and recognition, students engage in the cognitive work that chemistry requires: they calculate, compare, explain, and revise. Self-Determination Theory explains why this structure is motivating: it satisfies autonomy, competence, and relatedness. Constructivism and experiential learning explain why it is cognitively effective: students build understanding by solving problems and reflecting on their solutions. Cognitive Load Theory explains why leveling matters: it prevents working memory from being overwhelmed by a long, multi-step procedure. Behaviourism explains why reinforcement helps: it sustains effort through retry and revision. No single theory is sufficient; together, they explain why a strategy that is simultaneously challenging, social, and rewarding can produce both higher achievement and stronger motivation.

For practice, the study offers a feasible, low-cost model that does not depend on digital technology. Teachers can implement the Box Breaker Race with printed boxes, whiteboards, badges, and scoring sheets, using scenarios drawn from everyday life. School administrators can support this work by providing time for teachers to develop gamified materials, by recognising collaborative lesson design, and by ensuring that assessment includes both achievement and motivation. The enhanced strategy recommended here, with a brief procedural mini-lesson for the hardest competency, cooperative scoring that rewards helping, and a debriefing that makes the chemistry explicit, addresses the study's main weaknesses while preserving its strengths. Future research should replicate the study with a larger, randomised sample, include a delayed post-test to test retention, measure motivation before and after instruction, and collect qualitative data on how different learners experience competition and teamwork. If those replications confirm the present results, the Problem-Based Gamified Teaching Strategy deserves a place in the standard repertoire for teaching demanding chemistry topics, not as a replacement for direct instruction, but as a powerful complement that makes direct instruction more meaningful, memorable, and motivating.

REFERENCES

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. Learning and Instruction, 16(3), 183-198.
Alreshidi, N. A. K., & Lally, V. (2024). The effectiveness of training teachers in problem-based learning implementation on students' outcomes: A mixed-method study. Humanities and Social Sciences Communications, 11, 1137.
American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications. Educational Researcher, 35(6), 33-40.
Andal, G. Q., & Hermosa, J. P. (2024). The effectiveness of problem-based learning strategies in the creative thinking skills of Grade 12 students in Contemporary Philippine Arts from the Regions. International Journal of Multidisciplinary: Applied Business and Education Research, 5(2), 615-625.

Anderson, J. R., Reder, L. M., & Simon, H. A. (2000). Applications and misapplications of cognitive psychology to mathematics education. Texas Educational Review, 1(2), 29-49.
Azer, S. A. (2001). Problem-based learning: A critical review of its educational objectives and the rationale for its use. Saudi Medical Journal, 22(4), 299-305.
Bagay, M. C., Ursua, R. R. R., Abellera, M. A. A., Baldovino, R. J. G., Concepcion, R. A. P., Galapon, V. S., & Bautista, R. G. (2023). Problem-based learning in teaching science. Journal of Innovations in Teaching and Learning, 3(1).
Bell, J. (2012). Introducing problem-based learning as a learning strategy for master's students. ERIC.
BouJaoude, S., & Barakat, H. (2003). Students' problem-solving strategies in stoichiometry and their relationships to conceptual understanding and learning approaches. Electronic Journal of Science Education, 7(3), 1-42.
Chandrasegaran, A. L., Treagust, D. F., & Mocerino, M. (2007). The development of a two-tier multiple-choice diagnostic instrument for evaluating secondary school students' ability to describe and explain chemical reactions using multiple levels of representation. Chemistry Education Research and Practice, 8(3), 293-307.
Chans, G. M., & Portuguez Castro, M. (2021). Gamification as a strategy to increase motivation and engagement in higher education chemistry students. Computers, 10(10), 132.
Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum Associates.
Creswell, J. W., & Creswell, J. D. (2018). Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). SAGE Publications.
Csikszentmihalyi, M. (1990). Flow: The psychology of optimal experience. Harper & Row.
Department of Education. (2015). Policy guidelines on classroom assessment for the K to 12 Basic Education Program (DepEd Order No. 8, s. 2015). Republic of the Philippines.
Department of Education. (2025). Pilot implementation of the Strengthened Senior High School Curriculum, School Year 2025-2026 (DM No. 048, s. 2025). Republic of the Philippines.
Deunk, M. I., Smale-Jacobse, A. E., de Boer, H., Doolaard, S., & Bosker, R. J. (2018). Effective differentiation practices: A systematic review and meta-analysis. Educational Research Review, 24, 31-54.
Gisele, M. S., & Mukamana, S. (2024). Gamification in science education through artistic techniques. ResearchGate.
Hashim, N., Harun, O., Ariffin, N. A., & Chua, N. A. (2023). Gamification using board game approach in science education: A systematic review. Journal of Advanced Research in Applied Sciences and Engineering Technology, 33(3), 73-85.
Hou, H.-T. (2023). Diverse development and future challenges of game-based learning and gamified teaching research. Education Sciences, 13, 337.
Johnstone, A. H. (2010). You can't get there from here. Journal of Chemical Education, 87(1), 22-29.
Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work. Educational Psychologist, 41(2), 75-86.
Kulkarni, P., Prayag, G., Satish, Y. M., & Basavaraj, T. (2022). An empirical study on the impact of learning theory on gamification-based training programs. Organization Management Journal, 19(5), 170-188.
Li, M., Ma, S., & Shi, Y. (2023). Examining the effectiveness of gamification as a tool promoting teaching and learning in educational settings: A meta-analysis. Frontiers in Psychology, 14.
Marzano, R. J., Pickering, D., & Pollock, J. E. (2001). Classroom instruction that works. Association for Supervision and Curriculum Development.
Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), The Cambridge handbook of multimedia learning (2nd ed., pp. 43-71). Cambridge University Press.
Nieswandt, M. (2007). Student affect and conceptual understanding in learning chemistry. Journal of Research in Science Teaching, 44(7), 908-937.
Ostby, R. (2022). The impact of problem-based learning on students' critical thinking skills and peer relationships. Minnesota State University Moorhead.
Plass, J. L., Moreno, R., & Brunken, R. (Eds.). (2010). Cognitive load theory. Cambridge University Press.
Rosenshine, B. (2012). Principles of instruction: Research-based strategies that all teachers should know. American Educator, 36(1), 12-19.
Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist, 55(1), 68-78.
Smiderle, R., Rigo, S. J., Marques, L. B., et al. (2020). The impact of gamification on students' learning, engagement and behavior based on their personality traits. Smart Learning Environments, 7, 3.

Sousa, C., Rye, S., Sousa, M., et al. (2023). Playing at the school table: Systematic literature review of board, tabletop, and other analog game-based learning approaches. Frontiers in Psychology, 14.
Sweller, J. (2011). Cognitive load theory. In J. P. Mestre & B. H. Ross (Eds.), Psychology of learning and motivation (Vol. 55, pp. 37-76). Academic Press.
Sweller, J., van Merrienboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. Educational Psychology Review, 31(2), 261-292.
Taber, K. S. (2013). Revisiting the chemistry triplet. Chemistry Education Research and Practice, 14(2), 156-168.
Taspinar, B., Schmidt, W., & Schuhbauer, H. (2016). Gamification in education: A board game approach to knowledge acquisition. International Conference on Knowledge Management.
Towns, M. H., Raker, J. R., Becker, N., Harle, M., & Sutcliffe, J. (2012). The biochemistry tetrahedron and the development of the taxonomy of biochemistry external representations (TOBER). Chemistry Education Research and Practice, 13(3), 296-306.
Usta, E. (2023). Lifelong Learning Motivation Scale (LLMS): Validity and reliability study.
Walsh, N. (2020). Teaching percent composition, empirical and molecular formulas through the lens of aspirin. ChemEd X.
Wu, H.-K., Krajcik, J. S., & Soloway, E. (2001). Promoting understanding of chemical representations. Journal of Research in Science Teaching, 38(7), 821-842.
Zainuddin, Z., & Keumala, C. (2021). Gamification concept without digital platforms: A strategy for parents on motivating children to study at home during the COVID-19 pandemic. PEDAGOGIK: Jurnal Pendidikan, 8, 156-193.
Republic Act No. 10533. (2013). Enhanced Basic Education Act of 2013. Republic of the Philippines.