KIMIKA Volume 34, Number 2, pp. 87-113 (2023)
 ISSN 2508-0911
https://doi.org/10.26534/kimika.v34i2.87-113

Assessment of the Effectiveness of "Thermika," a Serious Game, in Teaching Thermochemistry

Miel Gabriel Lagdan*, Pauline Khyla D. Gabatin, Niño Sherwin M. Lumancas, Romalyn Balagtas-Arguelles, Toshio Akira Capintog

Caloocan City Science High School, P. Sevilla Street cor. 10th Avenue Grace Park West, Caloocan City, National Capital Region, Philippines 1406

*Author to whom correspondence should be addressed; email: miellagdan@gmail.com

ABSTRACT

With the emergence of novel advanced technologies, innovations in educational research have allowed the integration of learning into game-based approaches. As students view chemistry as a complicated subject, inadequate performance in the subject is observed. Game-based materials were shown to increase performance in their chosen subjects, and as such, this study aims to improve the performance of students in thermochemistry using a game-based approach called "Thermika." The evaluation of the app's suitability shows that the game is suitable, gaining a mean of 4.58. Analysis of the pre-test and post-test scores from participants who have participated in either the individual or group-based application of the intervention showed a significant difference ($p<0.001$) in their scores, which favored the post-test. This statistical significance is further supported by the approximately equal normalized gain score of 0.180 ($g=0.180$) of both groups and the respective effect sizes of 0.802 and 0.708 for the individual and group-based interventions, which ascertains the game's practical significance as having a large effect. As implied by the evaluations and tests done to determine the statistical and practical significance of the application of the intervention, "Thermika" proved to be a suitable and effective pedagogical tool for improving performance and understanding in thermochemistry.

Keywords: *serious game; chemistry; thermochemistry; game development; individual intervention; group intervention*

INTRODUCTION

Chemistry is considered a complex subject as the concepts that comprise chemical education today consist primarily of the abstract and thus can only be explained at the symbolic level, relying on making sense of teaching using invisible methods (Gabel, 1999; Ben-Zvi et al., 1987; Johnstone, 1991 as cited by Cardellini, 2012; Solikhin & Wijanarko, 2021).

With chemistry being a vast subject with varying branches, each containing abstract ideas, students do not just struggle with the subject but also find it uninteresting due to the usual outdated teaching methods, thick and wordy textbooks, absence of practical applications, inefficient resources, and lack of facilities (Richardson, 2022).

In the Philippines, secondary students' performance in this subject is considered low compared to international competency, as reasoning and analysis have been least emphasized in teaching the subject (Espinosa et al., 2014). This lack of adequate reasoning and analysis skills was also found in a study conducted by Seetee et al. (2016) when secondary students manifested issues regarding integrating science process skills in chemistry. Issues regarding reasoning and application can negatively affect the understanding and comprehension of chemistry topics (Nakthong et al., 2007, as cited by Seetee et al., 2016). In practice, this should be considered of great importance as "low science process skill will affect students' analytical thinking ability in chemistry learning achievement" (Irwanto et al., 2017, p. 5).

In the OECD Programme for International Student Assessment (PISA) 2018 results, the Philippines scored lower than the average OECD score compared to other ASEAN countries in science. The Philippines scored 357 points compared to the average score of 489 points, with the majority of the students who participated marked as having a proficiency level of 1 percent (Rosal et al., 2022; Department of Education - Cordillera Administrative Region, 2019).

In science education, the thermochemistry concept is vital and fulfills a developmental role. However, it is complex, and students have often misunderstood its concepts. Thermochemistry topics such as heat, temperature, enthalpy, and energy change are crucial in understanding all chemistry phenomena, making it very popular and complex for students to learn. The central dilemma in learning thermochemistry is the conceptual abstraction level. Students often encountered problems in thermochemistry, as shown by several conceptual research. For instance, the experiment performed by Mansoor Niaz shows that students cannot distinguish the concepts of heat and temperature (Rahmawati et al., 2021).

A study insists that students are still having difficulty understanding thermochemistry concepts. One of the considerable factors contributing to students' hardships is that thermochemistry is a complex subject since the students need to have prior knowledge in mathematics, physics, and chemistry to grasp the concepts of the subject. At the same time, the teaching conditions and the lack of focus also affected the student's performance in thermochemistry (Sokrat et al., 2014). A review emphasized inconsistencies in the scientific explanations, descriptions, understanding, and pedagogical approaches among the papers that affect the quality of learning chemistry (Bain et al., 2014).

With the emergence of new technologies, gamification has increased in education. With it, educators can provide gateways for an active, multisensory, and experimental environment for students to learn and develop critical thinking skills (Cheung & Ng, 2021). To aid the current situation in chemistry learning, games purposed for education called serious games (S.G.s) were created. It is implemented as "learners in serious games learned more, relative to those taught with conventional instruction methods, when the game was supplemented with other instruction methods, when multiple training sessions were involved, and when players worked in groups." (Garris et al., 2002, as cited by de Freitas, 2018, p. 78)

For instance, a study created a chemistry-centered game for engineering students taking a general chemistry class. Using the guidelines proposed by Gredler in his Handbook of Research for Educational Communications and Technology, the study generated a board game using chemistry-related questions to enrich student learning. This study showed that the game's effect on the student's construction of knowledge was positive as students were encouraged and aided to learn new concepts (Antunes et al., 2012). Also, an Android-based game, proposed as a pedagogical tool for chemistry, was created using Eclipse and the Android-development kit (ADK). Results indicated that the prototype meets all the criteria for usability achievement (Wan Ahmad & Abdul Rahman, 2014).

Concerning thermochemistry, current literature exists on creating an android-based learning media called Chemistry in Android (Chemdroid); it aims to improve students' achievement in thermochemistry. Their assessment results indicated that their mobile application is legible and valuable for students in various groups (Solikhin & Wijanarko, 2021). Likewise, Saraswati, Linda, and Herdini (Saraswati et al., 2019) developed an electronic magazine that focuses on the topics of thermochemistry—an Interactive E-Module Chemistry Magazine using Kvisoft Flipbook Maker with Plomp's model.

Following the successful attempts of multiple researchers using their serious games and the designed paradigm (Figure 1), the researchers tackled the issues regarding thermochemistry learning, mainly its abstract nature, using a serious game called "Thermika," that aimed to (a) improve performance in thermochemistry, and (b) develop a better understanding of its concepts.

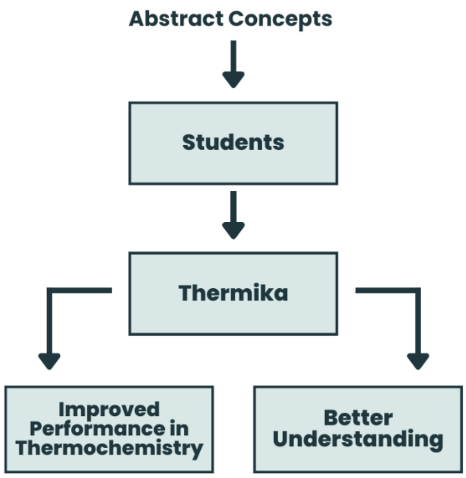

Figure 1. Research Paradigm

This study aimed to assess the suitability and effectiveness of "Thermika," a serious game, in teaching thermochemistry to senior high school students; specifically, the following research questions were explored: What is the suitability of the app as instructional material? (International Organization for Standardization, n.d.) in terms of Instructional Content, Functional Suitability, Performance Efficiency, and Usability. How effective is the app according to pre-test and post-test results by Individual Based Intervention and by Group Based Intervention.

The findings are meant to use the capacity of video games as an alternative to learning thermochemistry for students to enjoy studying, unaware that they are learning from the video game material they are using. The research can benefit the students by helping them to understand and apply the concepts of thermochemistry, aiding the teachers to expound upon thermochemistry and clarify misconceptions, assisting the school in fostering and inspiring the minds of future professionals, and, at last, being of service to the field of science education for making advancements in developing a thermochemistry-related serious game.

This research was limited to Caloocan City Science High School in the Philippines due to limited resources and time. Simultaneously, the topics covered by the game developed were limited to the most essential learning competencies (MELCs) set forth by the Philippine Department of Education for the General Chemistry 2 curriculum.

METHODS

The ADDIE model (Analysis, Design, Development, Implementation, and Evaluation), a learning model used by instructional designers and training developers to create effective learning experiences (DeBell, 2020), was used to develop the serious game (Figure 2). This model was appropriate to use as this methodology is extremely valuable for large and complex teaching designs (Bates & Bates, 2015) and is effective for creating professionally developed, efficient, and high-quality instructional material (ISFET, n.d.).

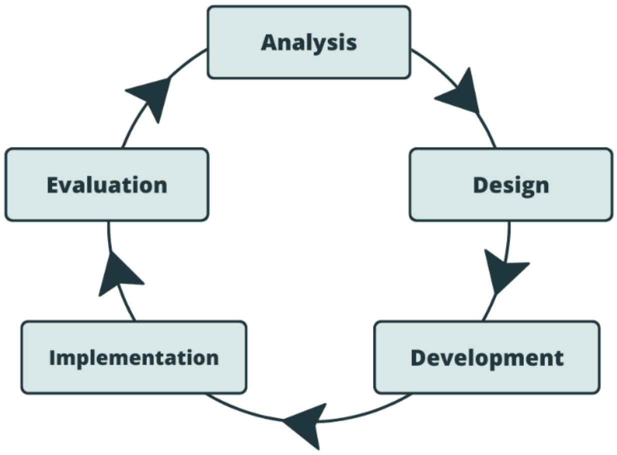

Figure 2. ADDIE Model of Instruction

Research Design. The researchers sought to find the game's suitability as a learning material and to test its effectiveness by the results of a pre-test and post-test. The researchers utilized a pre-experimental design using a one-group pre-test-post-test research design as groups already exist in the population. It also gives leverage to the researchers to test the dependent variable, which is the student's knowledge before and after the intervention, with the independent variable, which is the information given during the game implementation. The design was selected due to its simplicity in implementation and analysis. This ensures that the researchers can conduct the implementation phase of their study with little complexity (Cranmer, 2017, as cited in Allen, 2017).

PHASE 1: Development of "Thermika." *Analysis.* The teaching approaches to learning chemistry are often considered mundane and outdated, which often leads students to have difficulty learning and grasping the concepts of the subject (Richardson, 2022). Thermochemistry was selected as the subject of the study because it is a branch of chemistry that requires prior knowledge of basic chemistry, mathematics, and physics (Sokrat et al., 2014; Solikhin & Wijanarko, 2021). The materials utilized in the study were from Raymond Chang's Chemistry, Sixth Edition, as this was used to develop the Most Essential Learning Competencies (MELCs) provided by the Philippine Department of Education for the General Chemistry 2 curriculum (Table 1).

Table 1. Most Essential Learning Competencies of the General Chemistry 2 Curriculum

Content	Learning Competencies
Thermochemistry 1. Energy Changes in Chemical Reactions: exothermic and endothermic processes 2. First Law of Thermodynamics 3. Enthalpy of a Chemical Reaction 4. Calorimetry 5. Standard Enthalpy of Formation and Reaction Hess's Law	1. Explain the energy changes during chemical reactions. 2. Distinguish between exothermic and endothermic processes 3. Explain the first law of thermodynamics. 4. Explain the enthalpy of a reaction. 5. Write the thermochemical equation for a chemical reaction. 6. Calculate the change in enthalpy of a given reaction using Hess Law

Moreover, several graphics (Table 2) and "proportionality sliders" (Figure 3) were also integrated into the lesson prompts that allowed learners to have a visual representation of the indicated lesson and to help better understand the relationship of variables in concepts, respectively.

Table 2. Graphics Used in "Thermika"

Graphics Used	Source
Parts of a Thermodynamic System	What Is Thermodynamic System? - Open, Closed, & Isolated (With Examples). (n.d.). Law of Thermodynamics. https://lawofthermodynamicsinfo.com/what-is-thermodynamic-system/
Types of Thermodynamic Systems	Bagaria, M. (2023). Thermodynamic system and its type | Definition & Examples. Eigenplus. https://www.eigenplus.com/types-of-thermodynamic-system-definition-examples/
Exothermic vs. Endothermic Processes	Brearley, M. (2022). Exothermic and endothermic reactions. Online Learning College. https://online-learning-college.com/knowledge-hub/gcses/gcse-chemistry-help/exothermic-endothermic-reactions/
Parts of a Bomb Calorimeter	Bomb Calorimeter — Structure & Function - Expii. (n.d.). Expii. https://www.expii.com/t/bomb-calorimeter-structure-function-8685
Table of the Enthalpies of Formation of Common Elements and Compounds	Answered: TABLE 5.3 Standard Enthalpies of. . . | bartleby. (n.d.). https://www.bartleby.com/questions-and-answers/table-5.3-standard-enthalpies-of-formation-ah-at-298-k-dn-kjmol-substance-formula-dn-kmol-substance-/3681451c-422d-4c53-9994-457460a52b25
Graphical Representation of Hess's Law	Valqui, M. (2023). Hess's Law. ChemTalk. https://chemistrytalk.org/what-is-hesss-law/

Figure 3. Proportionality Sliders in the Game

Design and Development. In this stage, the results from the analysis were utilized in creating the concept of the game. The design of the game's mechanics, flow, and aesthetics have been deliberated and decided upon by the researchers using storyboards and flowcharts. The game comprised five chapters and an interactive minigame integrated with the lesson. Every chapter consisted of different thermochemistry topics, as stated in Table 1. Before the minigames, the lesson from the respective thermochemistry topics will prompt first to teach the players.

The game possesses a scoring, reward, and demerit system. The system permits that each minigame accomplished perfectly will give the player a maximum score of 100 points, and any incorrect answers will decrease the maximum score by 10. The highest score a player will receive, provided that the player has not repeated the same chapter, is 1300 points. Furthermore, the integrated minigames consist of real-life applications of the concepts taught to the players. It aims to provide supplementary knowledge for the concepts the players already know.

As for the game's development, the researchers used the Unity game engine. This game engine has a great community, a low entrance level, and a smooth learning curve. However, at the same time, it is strict and has many powerful features to write stable and effective codes. All this makes Unity technologies perfect for beginners and mature programmers (Titov, 2022).

The game's graphics, namely the map, control panel interface, and character sprites, were created in an online 2D-pixel generator site called Piskel, while the game's textual interface, like the dialogue box, lesson prompts, and splash screen, was generated with the use of Photoshop and Canva. However, the sprite assets for the pet feature for the reward system came from online sites such as DeviantArt and RPG Generator.

It can be seen in Figure 4, the map of the whole game. Figures 5 to 9 show the five minigames developed and arranged by chapters. Figure 4 shows the overall map of the virtual world of "Thermika." It contains all the places wherein the player will roam to explore various topics in thermochemistry. It features a power center, water treatment plant, train station, city hall, and neighborhood.

The first minigame (Figure 5) has two (2) parts. The first part is at the energy plant, where the player would have to classify the characteristics of different systems. The second part of the minigame will be held at the player suite, where the player will identify whether a process is exothermic or endothermic.

The second minigame (Figure 6) tests the player's knowledge of the First Law of Thermodynamics. A locked safe is presented to the player in the middle of the room. The code is obtained from the six (6) clues that were scattered around the area. These clues all revolve around the concept of the law of conservation of energy. Every clue has a corresponding number on the safe to avoid confusion on the placement of answers.

The third minigame (Figure 7) focuses on solving the enthalpy of a chemical reaction. This game presents two (2) problems for the players to solve on the enthalpy of a chemical reaction to fix and replace the broken pipes in the utility's facility. This minigame also aids in improving understanding of thermochemical equations and finding the amount of heat involved from a given thermochemical equation and the amount of substance.

The fourth minigame (Figure 8) tests the player's understanding of calorimetry by presenting three problems highlighting three important competencies, the relationship of heat capacity, mass, and specific heat, the formula for finding the amount of heat in Joules, and the proportionality of mass and heat. The problems are answered by making the user type the correct answer in the virtual keypad.

Figure 4. Map of the Game

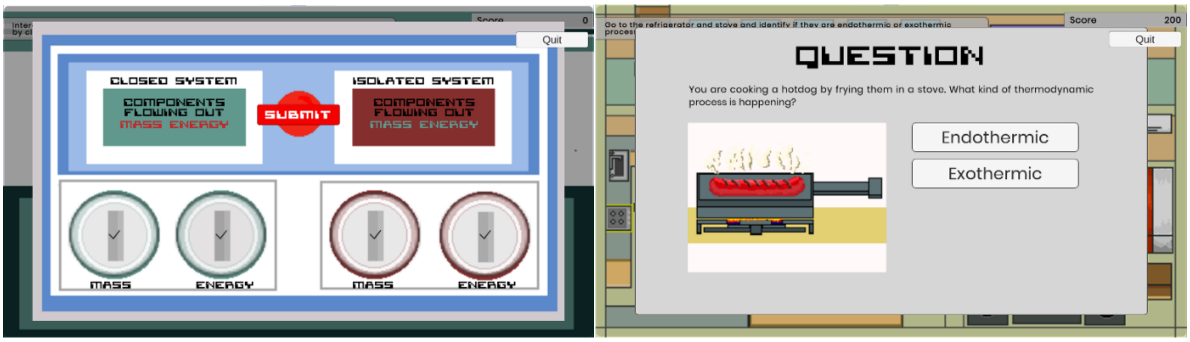

Figure 5. Chapter 1: Energy Changes in Chemical Reactions

Figure 6. Chapter 2: First Law of Thermodynamics

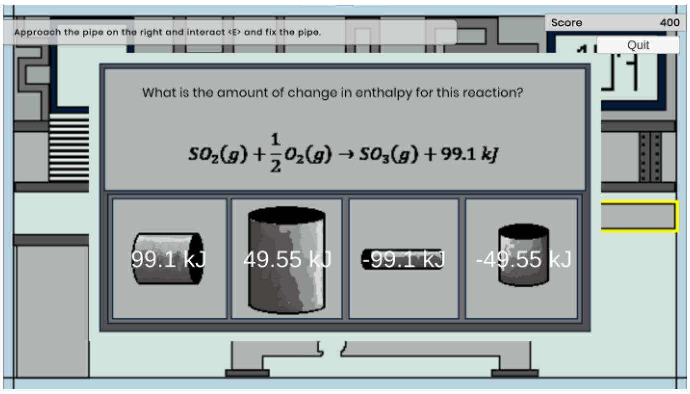

Figure 7. Chapter 3: Enthalpy of a Chemical Reaction

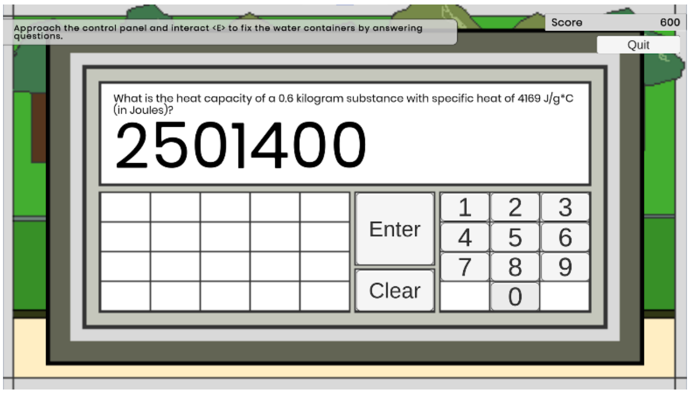

Figure 8. Chapter 4: Calorimetry

The fifth minigame (Figure 9) tests the players' knowledge of Standard Enthalpies of Formation and Hess's Law. The players needed to solve the two (2) problems using Standard Enthalpies of Formation and Hess's Law, respectively, to be able to move the trains from different stations. It does so by making the player solve the problems step-by-step, allowing the player to know the different properties of thermochemical equations.

Implementation and Evaluation. Selected professionals evaluated the intervention material based on their expertise to test its effectiveness. Furthermore, the experts were asked to provide a form of identification. It will act as proof of the authenticity and reliability of the evaluators.

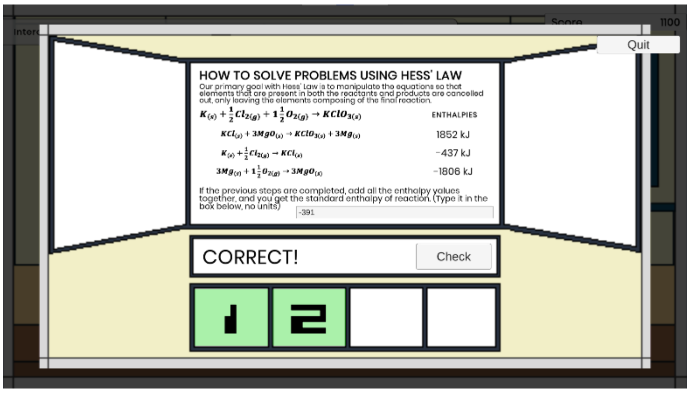

Figure 9. Chapter 5: Standard Enthalpies of Formation and Reaction Hess's Law

Questionnaires. A chemistry teacher and a master teacher evaluated the pre-test and post-test questionnaires to evaluate the face and content validities of the instrument. The validators focused on commenting on their suggestions as follows: the master teacher focused on face validity, while the chemistry teacher focused on content validity. After it underwent face and content validities, the questionnaires were used for the intervention.

To determine the reliability of the test, a test-retest reliability test was performed by applying the test to a group of people, then administering the same test to the same people five days later. The number of students with correct answers per item for both the pre-test and post-test were then analyzed for reliability using Pearson’s Correlation Coefficient. The results of the test are seen in Table 3.

Table 3. Test-Retest Reliability by Pearson’s Correlation Coefficient

	Pearson’s r	**p-value**
Pearson’s Correlation	0.813	< 0.001

As Pearson’s r is greater than 0.8, the test possesses good reliability, and is therefore, consistent in gathering data. This implies the questionnaire is sufficiently reliable to be used for the study.

Content evaluation. Five experts in the fields of chemistry and physics were chosen to assess the intervention material's content. The content validators were provided a DepEd Learning Resources Management and Development System (LRMDS) for Non-Print Materials evaluation sheet. An evaluation form focused on assessing the intervention material based on its content accuracy, material quality, visuals, and errors found. The results of the evaluation are shown in Table 4. The passing score of the content evaluation sheet is based on the DepEd standard given in the LRMDS considering that it varies with different criteria.

As shown in Table 4, three out of five of the evaluators failed the intervention material in the first evaluation with the same criterion: factor D. One also failed factors B and C. The evaluation results showed that the developed game's content quality, factor A, surpassed the chemistry experts' standard. It implies that the overall idea of the game was well executed and aligned with the thermochemistry topics according to the MELCs. However, factor B's instructional quality shows that the game was still lacking, and some components needed to be resolved as one of the evaluators was not satisfied. According to the evaluation sheet, the game lacks a challenge for the users, so it may not stimulate creativity and a good response from the players. They might not be able to have a sense of fun in learning. With that, the evaluators recommended imposing a reward system wherein the players can feel a sense of achievement and make it more enjoyable after passing through the chapters like in other games.

As an approach to the recommendation, the researchers thoroughly brainstormed with a reward system appropriate to the game. Until it ends with an item granting system rewards, the players can accumulate pets every chapter they have passed through. With this, it can maintain the interest of the players to finish the game even during the lull times and the lesson prompt (Wang & Sun, 2011). The game's technical quality, factor C, focuses on the game's visual and audio presentation. Validator 1 was not satisfied with the quality of the intervention material; thus, it needed improvement. According to the evaluation sheet, the validator remarks as poor one of the questions regarding the visual presentations in the game as being easy to interpret. It implies that some of the visual presentations can confuse players when playing the game. The other finding, factor D, consists of different errors as follows: factual, conceptual, grammatical, and other errors that can be found in the material. Even though some validators failed this criterion, it is easy for the researchers to amend it.

The researchers have solved the evaluators' comments before they re-evaluate it for the second time. Afterward, all the evaluators with failed criteria in evaluation became satisfied with adjustments done and passed all the criteria of LRMDS. Therefore, the intervention material has passed the content evaluation based on the content quality, instructional quality, technical quality, and other factors included in the LRMDS indicating that the intervention material can be utilized for the pilot testing.

Table 4. Scores from the Validators' LRMDS Evaluation Rating Sheet for Non-Print Materials

Validator	Factor A	Factor B	Factor C	Factor D	Other Comments
	Content Quality	Instr. Quality	Tech. Quality	Other Findings	
1	32	27*	38*	14*	Some content needs revisions. The reward system is lacking. The game must allow players to receive something meaningful.
1.1**	39	40	52	16	The game was revised according to the evaluator's comments.
2	33	30	39	14*	There are components, especially in the included questions, that need revisions. The term "heat" must be carefully used.
2.1**	33	30	39	16	The game was revised according to the evaluator's comments.
3	38	37	52	16	No other comments were made.
4	38	39	47	16	This project was well thought out and is a great way to get the attention of young students to be more interested in Science subjects.
5	40	40	52	15*	The voices did not match the dialogue text perfectly. Some questions and content of the lesson prompts need revision.
5.1**	40	40	52	16	The game was revised according to the evaluator's comments.

* Failed criterion, revisions were made accordingly.
** Re-evaluation scores.

PHASE 2: Pilot Testing. *Population and Sample.* The participants of this quantitative study were 84 grade 11 students, 15-17 years old with approximately the same number per gender (44 boys and 40 girls) at a legislated science high school enrolled in the STEM (Science, Technology, Engineering, and Mathematics) strand, which took GC2 subject in the second semester, inclusive of months January to May of 2023. The grade 11 STEM strand was divided into five sections, most of which were included in the study.

Instruments. To conduct the study, the researchers used the following instrument: the developed serious game in teaching thermochemistry, "Thermika" as the intervention material, and the pre-test and post-test questionnaires.

<u>Intervention Material.</u> Intervention material. The serious game was developed by the researchers using the Unity Engine. The game's content was aligned with the MELCs of General Chemistry 2 listed in Table 1. It consists of five (5) lessons and six (6) competencies.

<u>Pre-test and Post-test Questionnaire.</u> The validated questionnaire consisted of 25 randomly ordered questions. The questions involved measuring the students' prior knowledge regarding the concepts and applications of thermochemistry before the intervention. It also measures the gained knowledge after the intervention takes place.

Data Collection Procedure. To proceed with the implementation of the study, the permission of the school head was first sought. Furthermore, the General Chemistry teacher (author 4) assisted the researchers in implementing the intervention and data-gathering procedure. Afterward, the researchers oriented the participants in the implementation, precisely Grade 11 students. They are informed of the data collection flow from the pre-test until the post-test. The participants had a choice if they were going to participate or not.

The researchers used Messenger, Google Mail, Google Forms, itch.io, and Google Drive to conduct the study. The researchers administered the pre-test through ZipGrade in person with a pre-determined schedule. The researchers gave 30 minutes for the students to answer the pre-test. For the actual implementation of the game, the game published from the Unity Hub was compiled in a Google Drive and uploaded to itch.io for easier access for the participants. Moreover, the materials were emailed to participants who agreed to participate in the study. The participants were given five days to play the game.

As a compromise to those who were not able to play the game before the pre-determined date of the post-test, the participants were given the opportunity to play the game as a group by sharing their screen on a television screen for 30 minutes, as there was limited availability of devices capable of playing the game during the time allotted. Then, the post-test was administered through ZipGrade on a pre-determined date.

Ethical considerations. To ensure the study's ethical consideration, participants were required to submit parental consent forms through Google Forms. The Google Form consists of information the guardians need to know, such as the study's purpose. The parents were assured that their child's data would be kept strictly confidential and used only for research.

Statistical Analysis. The scores the participants gained from the pre-test and post-test were separated into individual and group-based players. Before the data analysis, the assumption of normality using the Shapiro-Wilk Normality Test was done since the validity of the results will rely on this test (Rosal et al., 2022). The results might differ for the two classifications; thus, the researchers assumed to use either Paired T-test for a normal distribution or Wilcoxon Signed Rank Test for a nonnormal distribution. Paired T-tests only require that the difference between the scores were normally distributed; thus, testing for equal variances was unnecessary (Xu et al., 2017). In calculating the effect size, it may be determined either by conducting Cohen's d or a rank biserial correlation. The interpretation of the effect size is displayed in Table 5, based from Cohen (1977).

Table 5. Effect Size Interpretation

Effect Size	Interpretation
0.2	Small Effect
0.5	Medium Effect
0.8	Large Effect

In determining gain, the normalized gain score (g), as introduced by Hake (1998) and mentioned by Coletta & Steinert (2020), is given by the formula:

$$g = \frac{\mu_{post}\% - \mu_{pre}\%}{100\% - \mu_{pre}\%}$$

Where $\mu_{post}\%$ is the percentage of the mean of the post-test over total number of items in the test and $\mu_{pre}\%$ is the percentage of the mean of the pre-test over the total number of items in the test.

RESULTS AND DISCUSSION

Suitability of the Application According to ISO 25010. This study evaluated the serious game's suitability as a digital instructional material using the Software Product Quality Model (ISO 25010). Utilizing a 5-Point Likert scale, the survey consisted of questions assessing the game based on the following parameters: a) Functional Suitability, b) Performance Efficiency, c) Usability, and d) Instructional Content. The evaluators suitable for assessing the game were I.T. experts. The researchers reached out to eight evaluators who have graduated and worked in the information technology industry or have degrees related to computer-based sciences. For the analysis of the evaluation from the ISO 25010, the scores obtained from the 5-Point Likert Scale were analyzed and computed for the mean and the standard deviation to determine the overall response for each criterion emphasized in the survey. Table 6 shows the interpretations of different ratings and value ranges for a 5-Point Likert Scale.

Table 6. Interpretation of a 5-Point Likert Scale for ISO 25010

Rating	Value Ranges	Interpretation
5	4.21 – 5.00	Strongly Agree
4	3.41 – 4.20	Agree
3	2.61 – 3.40	Acceptable
2	1.81 – 2.60	Disagree
1	1.00 – 1.80	Strongly Disagree

Table 7 depicts the values under the Instructional Content, particularly the Correctness of the Content, Relevance, and Engagement, with its corresponding mean scores of: 4.88, 4.75, and 4.63, accompanied by its respective standard deviations: 0.518, 0.463, and 0.518. The interpretation of the game corresponds with Strongly Agree, which suggests that the problems, concepts, and graphics found in the game were relevant, applicable, appealing, and accurate. Furthermore, the standard deviation values being close to each other also imply that the chances for random errors were minimized. The game had an above-average rating in all its parts. Despite the game receiving positive results, the evaluators suggested that the dialogues and lesson prompts must be more concise for the students to be more immersed and attentive and reduce the indecent of disinterest.

Table 7. The criteria and evaluation marks in Instructional Content as well as the Mean, Standard Deviation, and Mean Score Interpretation.

	Mean	Standard Deviation	Interpretation
Instructional Content	4.75	0.445	Strongly Agree
Correctness Of the Content	4.88	0.518	Strongly Agree
Relevance	4.75	0.463	Strongly Agree
Engagement	4.63	0.518	Strongly Agree

The game has accumulated a mean score of 4.75 and a standard deviation of 0.445 in Instructional Content, which implies that the evaluators of the game agree that it is a viable instructional material in imparting grade 11 STEM students about the concepts of Thermochemistry according to the MELCs of General Chemistry 2 listed in Table 1.

Table 8 depicts the values that compose the criterion for Functional Stability. It consists of Functional Appropriateness and Functional Correctness with corresponding mean scores of: 4.75 and 4.25, respectively, with corresponding values for standard deviations: 0.463 and 0.707. The collected mean values are interpreted as Strongly Agree, implying that the game is proper for its intended demographic. The values of the standard deviations are relatively close to each other, implying that the chances of random errors are minimized. Furthermore, the game has accumulated a mean of 4.50 with a standard deviation of 0.707 in Functional suitability, which means that the game's evaluators agree that it is proper and not overwhelming for its target demographic as well as a beneficial teaching aid in learning Thermochemistry.

Table 8. The criteria and evaluation marks in Functional Suitability as well as the Mean, Standard Deviation, and Mean Score Interpretation.

	Mean	Standard Deviation	Interpretation
Functional Suitability	4.50	0.585	Strongly Agree
Functional Appropriate	4.75	0.463	Strongly Agree
Functional Correctness	4.25	0.585	Strongly Agree

Table 9 depicts the values under Performance Efficiency, which is solely composed of Resource Utilization with a mean score of 4.63 with a standard deviation of 1.061. The mean interpretation of Strongly Agree implies that the evaluators agree that the game can perform as intended. Since Performance efficiency only consists of a single component, the mean and standard deviation for Resource Utilization would be the same. Due to this implication, there is a chance of a random error to occur. This means there might be a chance that some of the evaluators did not experience any bugs or glitches that were made through the final version game.

Table 9. The criteria and evaluation marks in Performance Efficiency as well as the Mean, Standard Deviation, and Mean Score Interpretation.

	Mean	Standard Deviation	Interpretation
Performance Efficiency	4.63	1.061	Strongly Agree
Resource Utilization	4.63	1.061	Strongly Agree

Table 10 depicts the values under Usability composed of Operability, Learnability, and User Inference Aesthetics with the respective mean scores of: 4.50, 4.63, and 4.30 and their corresponding values of standard deviations: 0.756, 0.744, and 0.463, respectively. Though the game had a positive response from the evaluators, some gave their opinions on the aesthetic features that would have made it much more pleasing to the players. The game gathered a mean of 4.46 with a standard deviation of 0.654 with a corresponding interpreted value of Strongly Agree. The interpretation implies that the game has easy-to-follow instructions, encourages one to learn thermochemistry, and has a pleasing interface. With the standard deviation values being close together, it implies that it has minimized the chances for random error to occur.

Table 10. The criteria and evaluation marks in Usability as well as the Mean, Standard Deviation, and Mean Score Interpretation.

	Mean	Standard Deviation	Interpretation
Usability	4.46	0.654	Strongly Agree
Operability	4.50	0.756	Strongly Agree
Learnability	4.63	0.744	Strongly Agree
User Interface Aesthetics	4.30	0.463	Strongly Agree

Table 11 summarizes the mean scores and standard deviations of all the presented categories for ISO 25010: Instructional Content, Functional Suitability, Performance Efficiency, and Usability. The overall mean of all the abovementioned combined factors is 4.58, with a standard deviation of 0.686. The overall mean and the factors that it consists of were interpreted as Strongly Agree, and the standard deviation value was below zero, implying that the game is a successful web-based desktop teaching material. A similar study was conducted by Solikhin & Wijanarko (2021) which showed that educational technology piques the interest of students in educational technology and enables a better assessment of the subject matter, builds knowledge about the use of game-based learning, enables a wider variety of learning Thermochemistry, develops their metacognitive abilities, and promotes chemistry to students.

Table 11. Mean, Standard Deviation, and Mean Score Interpretation Marks in Each Category

	Mean	Standard Deviation	Interpretation
Instructional Content	4.58	0.445	Strongly Agree
Functional Suitability	4.50	0.585	Strongly Agree
Performance Efficiency	4.63	1.061	Strongly Agree
Usability	4.46	0.654	Strongly Agree
Overall	4.58	0.686	Strongly Agree

Effectiveness of the Application According to Pre-Test and Post-Test Assessments. This section discussed the results of the assessments in determining the effectivity of "Thermika" in improving the participants' thermochemistry knowledge. It determines how the participants performed in the assessments before and after the intervention was applied individually and as a group, as proposed for analyzing pre-test and post-test scores by York (2017).

Intervention Applied Individually. Test for Normality and Equal Variance. To determine the appropriate statistical test for the correct interpretation of the given data, it first underwent a test for normality. The Shapiro-Wilk Normality Test (Shapiro & Wilk, 1965) was used, with a sample size of 46 ($n = 46$), the Shapiro-Wilk Normality Test deemed appropriate for such a small sample size, defined as a sample size less than 50 ($n < 50$), as it is more reliable in detecting nonnormality (Mishra et al., 2019). As the data set was paired, the difference between the pre-test and post-test scores was used to determine the normality. The test showed no evidence of nonnormality, as presented in Table 12 Figure 10, as the significance (p-value) was greater than 0.05 ($p > 0.05$), which fails to reject the null hypothesis that the data set is normally distributed.

Table 12. Test for Normality – Shapiro-Wilk Normality Test

	Statistic (W)	df	Significance (p)
Difference	0.959	46	0.107

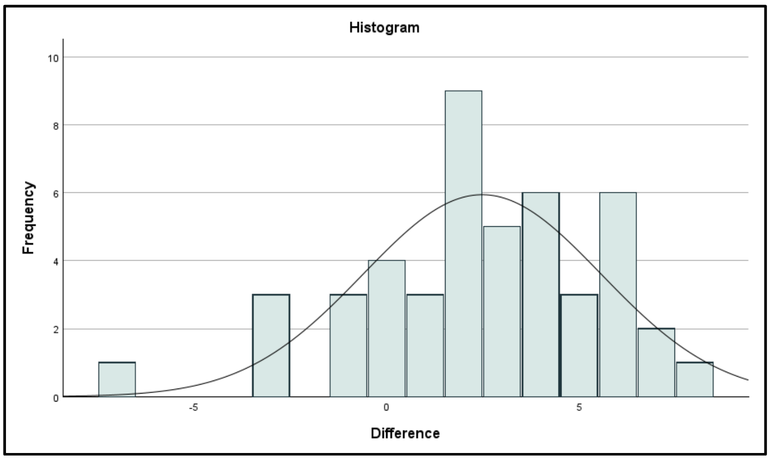

Figure 10. Frequency Distribution Histogram of Difference in Pre-Test and Post-Test Scores

Given the normality of the data set as proven by the Shapiro-Wilk Normality Test, the appropriate statistical test to use was a paired sample t-test. No test for equal variances was conducted as the paired sample t-test does not require the assumption of equal variances (Xu et al., 2017).

<u>Student's Knowledge Before Intervention.</u> To measure the knowledge possessed by the participants in thermochemistry prior to the application of the intervention, a pre-test was administered. The mean, median, standard deviation, and minimum and maximum are found in Table 13.

Table 13. Pre-Test Results

	N	Mean	Median	Standard Deviation	Minimum	Maximum
Pre-Test	46	11.8	12.0	3.47	2	18

With a mean of 11.8 ($\bar{x} = 11.8$), this implies that the participants possess prior knowledge of thermochemistry. This indication is helpful as it was used as a baseline in determining improvement in knowledge. Afterward, the participants were given the game to play for five days, then after; a post-test was administered.

<u>Students' Knowledge After Intervention.</u> After using the intervention, the students took a post-test to assess their current knowledge in the unit. Table 14 presents the post-test results, including the mean, median, standard deviation, and minimum and maximum.

Table 14. Post-Test Results

	N	Mean	Median	Standard Deviation	Minimum	Maximum
Post-Test	46	14.2	14.0	3.20	9	22

With a mean of 14.2 ($\bar{x} = 14.2$), the post-test indicated a noticeable difference against the previously-mentioned pre-test scores, implying the intervention may have improved the participants' knowledge.

To test if this difference is significant, a paired t-test was performed. With a statistical significance level of 5% ($p = 0.05$), the common significance level used in educational research (Carver, 1978), the t-test results and normalized gain score (g) are presented in Table 15.

Table 15. Paired T-Test Results

Descriptives		Confidence Level (95%)		Significance			
Mean	**Std. Deviation**	**Lower**	**Upper**	**t**	**df**	**p**	**g**
-2.48	3.09	-3.40	-1.56	-5.44	45	< 0.001	0.182

Table 15 explains that there is a significant difference between the pre-test and post-test scores, as indicated by the p-value ($p < 0.001$), which provides enough evidence to reject the claim that the difference in pre-test and post-test scores is of random chance. These results provide enough evidence that applying the intervention, the game "Thermika," helped improve students' knowledge of thermochemistry.

The pre-test scores indicate a reasonably low score compared to the post-test scores. After the application of the serious game, the participants were able to understand thermochemistry concepts better given the addition of a "fun" factor (application of a reward system, integration of lessons in real-life scenarios, and addition of games to encourage critical thinking), which allowed the students to become more attentive to the lessons integrated into the intervention, as proven by the post-test scores.

21st-century technology allowed the integration of learning into multiple print and non-print media, and with the increasing growth of digital devices at home, the game "Thermika" allowed students to learn independently with the assistance of their teachers and instructors. This statistically significant difference in the pre-test and post-test scores and normalized gain score of 0.182 signifies that the intervention effectively improves performance and understanding of thermochemistry. This may also be proven by calculating the effect size using Cohen's d (Cohen, 1977). As presented in Table 16, the effect size can help better understand the practical significance, which is the significance in terms of standard deviations (York, 2017).

Table 16. Effect Size using Cohen's d

	Standardizer	**Point Estimate**	**Significance (95%)**	
			Lower	**Upper**
Cohen's d	3.089	-0.802	-1.132	-0.466

The effect size of 0.802, disregarding the negative sign, implies that the difference in pre-test and post-test is also practically significant, as an effect size equivalent to 0.8 is considered to impart a large effect. This means the post-test scores are almost one standard deviation better than the pre-test scores. In a real-world setting, this implies that the game is also effective in practical applications (e.g., in education).

<u>Students' Achievement Per Competency.</u> In terms of identifying improvement in student achievement per competency, the administered test consists of questions that measure the student's mastery in each competency listed in Table 1. Students' achievement per competency was determined by computing the frequency of correct students per item and its corresponding mean per question in the competency.

Table 17 indicates that most of the competencies listed in Table 1, except for competency seven, show increased frequency and mean of participants who answered correctly. However, the decrease in frequency and mean in competency two is observed, distinguishing exothermic and endothermic processes. This decrease was also observed in a study by Rosal et al. (2022). This implies an observed significant difference in the majority of the competencies that may be attributed to the application of the intervention.

Table 17. Frequency of Students and Mean of Pre-test and Post-test Result per Learning Competency

Learning Competency	Item No.	Pre-Test		Post-Test	
		Frequency	Mean	Frequency	Mean
1. Explain the energy changes during a chemical reaction.	2,18,19	79	26.33	108	36.00
2. Distinguish between exothermic and endothermic processes.	5,11	55	27.50	53	26.50
3. Explain the first law of thermodynamics.	6,16,21,22,23	88	17.60	98	19.60
4. Explain the enthalpy of a reaction.	12,14,17	113	37.67	119	39.67
5. Write the thermochemical equation for a chemical reaction.	10,24	33	16.50	45	22.50
6. Calculate the change in enthalpy of a given reaction using Hess' Law.	1,4,7,9,15	103	20.60	120	24.00
7. Explain heat changes using calorimetry. *	3,8,13,20,25	70	14.00	112	22.40

* This is a competency not listed in the MELCS in General Chemistry 2; however, the unit includes the lesson.

Category of Student Improvement. The students were then categorized according to their improvement of the mastery of the competencies in Table 1 into five categories, as proposed by Shukor et al. (2015).

Table 18 summarizes the score intervals for each category. The scores of the pre-test and post-test were categorized using the aforementioned, and the two categories were then compared to determine the final category of the students given in Table 19. Additionally, Table 19 gives the final categories categorized by the student's level of improvement toward the subject.

Table 18. Score intervals for each category of mastery

Scores	Category
21-25	High-Achiever, H-A
16-20	Medium-Achiever, M-A
11-15	Low-Achiever, L-A
6-10	Weak, W
0-5	Poor, P

Table 20 and Table 21 show the number of students in each mastery category and the number of participants' whom improved as indicated by the interpretations in Table 19, respectively.

Indicated by the change of frequency of the number of students per mastery category between the pre-test and post-test, it can be observed that the application of the intervention material may have contributed to helping students improve within their respective categories. It is observed that the numbers within the P (poor) and W (weak) categories have decreased, allowing an increase within the L-A (low achievement), M-A (medium achievement), and H-A (high achievement) categories, which, again, may be contributed to the application of the intervention material.

Table 19. Student's improvement categories

Category [Pre-Test] (i)	Category [Post-Test] (f)	Final Category	Symbol
P_i	W_f	+1 Improvement	P1
W_i	L-A_f	+1 Improvement	P1
L-A_i	M-A_f	+1 Improvement	P1
M-A_i	H-A_f	+1 Improvement	P1
P_i	P_f	Static	S
W_i	W_f	Static	S
L-A_i	L-A_f	Static	S
M-A_i	M-A_f	Static	S
H-A_i	H-A_f	Static	S
P_i	L-A_f / M-A_f / H-A_f	+2/+3/+4 Improvement	P2/P3/P4
W_i	M-A_f / H-A_f	+2/+3 Improvement	P2/P3
L_i	H-A_f	+2 Improvement	P2

Table 20. Frequency of pre-test and post-test scores per each mastery category

Category	Pre-Test	Post-Test
P	2	0
W	15	5
L-A	22	29
M-A	7	9
H-A	0	3

In terms of improvement, Table 21 displays how many students increased a category or retained after taking the post-test. Observed is majority of the individual participants increased a category after the post-test, while the rest retained the same ranking. This may imply that the use of the intervention material made a substantial difference in the scores of the participants that the change in scores may be considered as a noticeable improvement. However, also observed is the retained category which possesses a considerable number of students, which may imply that the intervention also caused a change too miniscule or non-existent to be considered as noticeable improvement. Present was also those who have not shown an improved or retained status, as their rank decreased.

Table 21. Frequency of students belonging to each improvement category

Improvement Category	Interpretation	Frequency
P1	+1 Improvement	20
P2	+2 Improvement	3
P3	+3 Improvement	0
P4	+4 Improvement	0
S	Static	20

Intervention Applied as a Group. This part of the analysis proved imperative to determine the difference between the application of the intervention individually or as a group.

<u>Test for Normality and Equal Variance.</u> The Shapiro-Wilk Normality Test (Shapiro & Wilk, 1965) was once again used to determine the normality of the data set to identify the appropriate statistical method to use.

With a sample size of 38 ($n = 38$), the differences in pre-test and post-test scores for each case were tested for their normality. The test shows the presence of nonnormality, as presented in Table 22, Figure 11, as the significance (p-value) is less than 0.05 ($p < 0.05$), which rejects the null hypothesis that the data set is normally distributed.

Table 22. Test for Normality – Shapiro-Wilk Normality Test

	Statistic (W)	df	Significance (p)
Difference	0.910	38	0.005

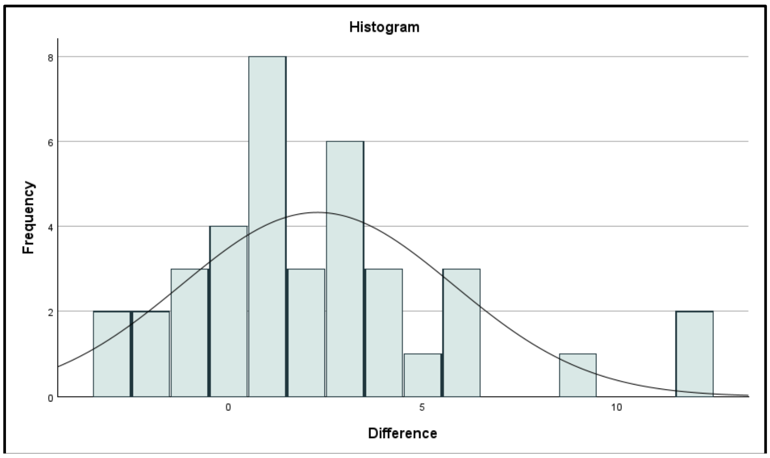

Figure 11. Frequency Distribution Histogram of Difference in Pre-Test and Post-Test Scores

Given the nonnormality of the data set, as proven by the Shapiro-Wilk Normality Test, the appropriate statistical test to use is Wilcoxon's signed rank test (Wilcoxon, 1992). This is because the test requires no assumption of normality and is used for paired samples (Woolson, 2008).

<u>Students' Knowledge Before Intervention.</u> As with the individual application, a pre-test was administered to gauge the participants' background knowledge in thermochemistry. The mean, median, standard deviation, and minimum and maximum are presented in Table 23.

Table 23. Pre-Test Results

	N	Mean	Median	Standard Deviation	Minimum	Maximum
Pre-Test	38	12.2	12.0	2.78	6	17

With a mean of 12.21 ($\bar{x} = 12.2$), the pre-test indicates that the participants who have played the game as a group have a background knowledge in thermochemistry slightly higher than those who have played the game individually ($\bar{x} = 11.8$). As mentioned in the data collection procedure, some of the participants were not able to play the game individually, hence they were given the opportunity to play the game as a group by sharing an instance of the game on a television screen, and then a post-test was conducted.

<u>Students' Knowledge After Intervention.</u> A post-test was conducted on participants who used the intervention as a group. The results of their post-test are presented in Table 24, along with the mean, median, standard deviation, and minimum and maximum.

Table 24. Post-Test Results

	N	Mean	Median	Standard Deviation	Minimum	Maximum
Post-Test	38	14.5	15.0	3.36	8	20

A similar observation to the first data set was also observed in the second, with an increase in post-test scores compared to the pre-test. With a mean of 14.50 ($\bar{x} = 14.50$); this increase may also be attributed to the application of the intervention.

As such, Wilcoxon's signed rank test was performed (Table 25) to test the significance of this difference in mean scores. Using a significance level of 5% ($p = 0.05$), the test was conducted, and the test results and normalized gain score (g) are shown in Table 26.

Table 25. Wilcoxon's Signed Rank Test - Ranks

		N	Mean Rank	Sum of Ranks
Pre-Test – Post-Test	**Negative Ranks**	7	12.4	87*
	Positive Ranks	27	18.8	508
	Ties	4		
	Total	38		

* The W statistic used as the smaller sum of the ranks is used.

Table 26. Wilcoxon's Signed Rank Test – Results

	Z*	Asymptotic Significance	g
Pre-Test – Post-Test	-3.621	< 0.001	0.180

* This is based on negative ranks.

Table 26, similar to Table 15, explains that the difference between the pre-test and post-test scores are considered statistically significant as the p-value taken from the conducted Wilcoxon's signed rank test ($p < 0.001$) provides sufficient evidence to reject the null hypothesis, implying the difference is not of random chance. Given the significant difference in scores and the normalized gain score of 0.180 as a result of the application of the intervention, this further fortifies the effectiveness of the game "Thermika". The results give insight that the game effectively improved the thermochemistry knowledge and understanding even if applied in a group setting, making it possible for use in a classroom during discussions.

In addition to the "fun" factor mentioned previously, this increase in post-test scores in a group setting may also contribute to the sense of competition and collaboration observed when the participants were playing the game, as (Subhash & Cudney, 2018) proposed.

In reporting the effect size for this data set, Cohen's d cannot be used as it operates under the assumption that the data set is normal (Cohen, 1977). Instead, as Mangiafico (2016) and King et al. (2018) recommended, a rank biserial correlation test should be used to find the effect size. The result of the test is shown in Table 27.

Table 27. Rank Biserial Correlation Results

	Effect Size
Rank Biserial Correlation	-0.708

Utilizing the same value ranges used to interpret Cohen's d, an effect size of 0.708 interprets that the game, applied in a group setting, also imparts a large effect. This statistic helps improve understanding of the game's practical significance, as having a large effect size implies that game helped to create a meaningful difference in the pre-test and post-test scores.

Students' Achievement Per Competency. As done with the first group, an analysis regarding the mastery of competencies listed in Table 1, given the frequency of correct students per item, provided in Table 28.

Table 28. Frequency of Students and Mean of Pre-test and Post-test Result per Learning Competency

Learning Competency	Item No.	Pre-Test		Post-Test	
		Frequency	Mean	Frequency	Mean
1. Explain the energy changes during a chemical reaction.	2,18,19	70	23.33	91	30.33
2. Distinguish between exothermic and endothermic processes.	5,11	46	23.00	46	23.00
3. Explain the first law of thermodynamics.	6,16,21,22,23	78	15.60	82	16.40
4. Explain the enthalpy of a reaction.	12,14,17	84	28.00	93	31.00
5. Write the thermochemical equation for a chemical reaction.	10,24	33	16.50	48	24.00
6. Calculate the change in enthalpy of a given reaction using Hess' Law.	1,4,7,9,15	81	16.20	102	20.40
7. Explain heat changes using calorimetry. *	3,8,13,20,25	72	14.40	89	17.80

* This is a competency not listed in the MELCS in General Chemistry 2; however, the lesson is included in the unit.

As seen from the frequency and mean of each competency in Table 28, the same event is observed as in Table 17. The majority of the competencies showed an increase in their respective frequencies and means, except for competency two. Given the consistency of this event in both the individual and group-based intervention, there is reason to speculate that the explanation of competency two in the game was not thorough enough to create any meaningful difference in the pre-test and post-test scores of the participants.

Also, by the consistency of results, the data presented in Table 17 and Table 28 suggests that the use of the game helped to improve student achievement and performance, as proven evident by the increase of the mean of the majority of the competencies included in the thermochemistry unit. This is consistent with studies conducted by Rosal et al. (2022) and Cordova et al. (2019), where the application of an electronic strategic intervention material (E-SIM) improved mastery in competencies under a specific subject and topic.

Category of Student Improvement. As done with the individual-based intervention, the students who participated in the group-based intervention were also categorized according to their level of mastery before and after the application of the intervention material (Table 29).

Table 29. Frequency of pre-test and post-test scores per each mastery category

Category	Pre-Test	Post-Test
P	0	0
W	11	6
L-A	22	14
M-A	5	18
H-A	0	0

Table 29 displays the same event observed in Table 20. After the administration of the intervention and post-test, observations on a change in the participants' categories, specifically, a change favoring an increase of rank, as the W (weak) and L-A (low achievement) categories' frequency of students decreased, while the M-A (medium achievement) category increased.

Table 30. Frequency of students belonging to each improvement category

Improvement Category	Interpretation	Frequency
P1	+1 Improvement	12
P2	+2 Improvement	4
P3	+3 Improvement	0
P4	+4 Improvement	0
S	Static	20

Observed in the individual application of the intervention, students who improved, either in the P1 or P2 category, outweighed in the S category (Figure 30). However, with the group intervention, the opposite is shown. The majority of the students retained their current rank, while only 16 improved in their respective categories. This may imply that, although it can improve the mastery of the students in thermochemistry topics, application of the intervention material in a group setting may not be as effective as compared to the individual application of the intervention. Thus, in terms of improvement in a group-based intervention, Thermika effectively improved students' mastery of thermochemistry but was also able to incur change too little to become noticeable. Also, a decrease in the category was observed, which may be due to a decrease in the test scores observed in some of the participants' tests.

The implementation of a group-based approach to learning thermochemistry would assist students who are not that competent in the subject in participating in active discussions with their peers. This approach would enable them to absorb and comprehend information with minimal mental effort, as observed in the study by Kirschner et al. in 2009. The study demonstrates that learning through a group develops an individual's cognitive load, resulting in a higher processing capacity for complex tasks.

In evaluating whether the application of the intervention individually or as a group made such a significant difference, the normalized gain scores of the two groups were compared. As their normalized gain scores are similar, 0.182 for individual application ($g = 0.182$) and 0.180 for the group application ($g = 0.180$), there is no such significant difference present in gain, and as such, considering the effect the intervention indifferent in applying individually or by a group.

With regards to the research objective, in determining the suitability of the application as an instructional material, the evaluation results, as displayed in Table 6 to Table 10, indicate that "Thermika" is suitable according to its instructional content, functional suitability, performance efficiency, and usability. With an overall mean score of 4.58 out of 5, the game is considered within the standards of a usable game fit for students to use.

In addition, it provided evidence of the statistical and practical significance of the game in both the individual and group application, as proven by the significant difference in pre-test and post-test scores and effect sizes. The serious game, "Thermika," effectively improves the knowledge and understanding of thermochemistry of grade 11 students. These findings are consistent with studies conducted by Astiningsih & Partana (2020), Solikhin & Wijanarko (2021), and Fitriyana et al. (2020), as the application of a game-based instructional material to student participants was observed to increase mean post-test scores in comparison to the mean pre-test scores.

These findings indicate and prove that "Thermika" is suitable and effective for becoming an independent pedagogical tool for students. These findings are significant to research involving game-based learning, as it enables the exploration of options for effective learning methods that may vary for each student.

This study was limited to grade 11 students in a legislated science high school in the Philippines. As such, the results of this study are not generalizable with other grade levels and locations. Further research should be explored regarding the applicability of the game in varying grade levels. Also, with only a small sample of the individual ($n = 46$) and group application ($n = 38$), larger sample sizes should be used in further research regarding "Thermika."

CONCLUSIONS

The game's ratings from eight I.T. experts have passed all the evaluation criteria for the game's suitability as an instructional material. The game's Instructional Content suggests that the game is factually correct, relevant, and engaging; Functional Appropriateness interprets the game as suitable and functional; Performance Efficiency interprets the game as playable. Operability shows that the game is easy and enjoyable, encourages people to present and learn, and has excellent interface aesthetics. Most of the scores from the evaluation criteria were close to the highest rating, 5, and all criteria were interpreted as Strongly Agree. Thus, the overall remark of the game "Thermika" is Strongly Agree with a mean score of 4.58 and a standard deviation of 0.686, where it indicates that the evaluators firmly agree that the application is suitable. It shows that "Thermika" is suitable as an app for teaching thermochemistry regarding its concepts and technicalities.

In addition, predicated on the tests conducted to see if students will manifest improved performance in thermochemistry given the application of the intervention, the observed data showed an increase in student performance in the topic with a normalized gain score of 0.180. The effectiveness of the game as an instructional material showed that the difference between the pre-test and post-test scores was indeed significant, with a p-value less than 0.001 ($p < 0.001$), favoring the post-test scores. As for the practical significance, the 0.8 effect size of the intervention material considers the intervention material to impart a large and practical effect on performance. Thus, it can be concluded that "Thermika" is an effective tool for improving thermochemistry performance and understanding.

Further studies are recommended to explore the versatility of the application across different platforms. The intervention material is bound only to the desktop platform. With this, research is recommended on optimizing the intervention material for the mobile platform. In addition, the researchers would also suggest future studies that involve the testing and implementation of pedagogical materials in students to consider having a control group to enforce the validity and reliability. As well as to observe if there is a significant change between the samples that were given the intervention and those that were not given any.

REFERENCES

Allen M. The SAGE Encyclopedia of Communication Research Methods [Internet]. 2455 Teller Road, Thousand Oaks California 91320: SAGE Publications, Inc; 2017 [cited 2023 May 28]. Available from: https://methods.sagepub.com/reference/the-sage-encyclopedia-of-communication-research-methods

Antunes M, Pacheco MAR, Giovanela M. Design and implementation of an educational game for teaching Chemistry in higher education. J Chem Educ. 2012 Jan;89(4):517–521. https://doi.org/10.1021/ed2003077

Astiningsih AD, Partana CF. Using Android media for Chemistry learning construction of motivation and metacognition ability. Int J Instruction. 2020 Jan;13(1):279–294. Available from https://files.eric.ed.gov/fulltext/EJ1239195.pdf

Bain K, Moon A, Mack MR. Towns MH. A review of research on the teaching and learning of thermodynamics at the university level. Chem Educ Res Pract. 2014 Feb;15(3):320–335. https://doi.org/10.1039/C4RP00011K

Bates AW (Tony), Bates AW. 4.3 The ADDIE model. 2015 Apr 5 [cited 2023 Jan 11]; Available from: https://opentextbc.ca/teachinginadigitalage/chapter/6-5-the-addie-model/

Ben-Zvi R, Silberstein J, Eylon RS. Students' visualization of a chemical reaction. Educ Chem. 1987 July;24(4):117–120. Available from https://weizmann.elsevierpure.com/ws/portalfiles/portal/112548368/rm_EducationInChemistry_StudentsVisualisation_PV1987.pdf

Cardellini L. Chemistry: Why the subject is difficult? Educación química. 2012 May;23(2):305–310. https://doi.org/10.1016/S0187-893X(17)30158-1

Carver R. The case against statistical significance testing. Harvard Educational Review. 1978 Sep 1;48(3):378–99.

Cheung SY, Ng KY. Application of the educational game to enhance student learning. Frontiers Educ. 2021 March;6:2021 Available from: https://www.frontiersin.org/articles/10.3389/feduc.2021.623793

Cohen J. Statistical power analysis for the behavioral sciences. Rev. ed. New York: Academic Press; 1977. 474 p.

Coletta VP, Steinert JJ. Why normalized gain should continue to be used in analyzing preinstruction and postinstruction scores on concept inventories. Phys Rev Phys Educ Res. 2020 Feb;16(1):010108. https://doi.org/10.1103/PhysRevPhysEducRes.16.010108

Cordova RC, Medina JGD, Ramos TR, Alejo AR. Effectiveness of Competency-Based Strategic Intervention Materials in English. DLSU Research Congress 2019. Available from https://www.dlsu.edu.ph/wp-content/uploads/pdf/conferences/research-congress-proceedings/2019/lli-II-019.pdf

Cranmer GA. One-Group Pretest–Posttest Design. In: Allen M, editor The SAGE Encyclopedia of Communication Research Methods. SAGE Publications, Inc. 2017. https://doi.org/10.4135/9781483381411

de Freitas S. Are games effective learning tools? A review of educational games. J Educ Technol Soc. 2018 April;21(2):74–84. Available from https://www.jstor.org/stable/26388380

DeBell A. What is the ADDIE model of instructional design? [Internet]. Water Bear Learning. 2020 [cited 2022 Dec 23]. Available from: https://waterbearlearning.com/addie-model-instructional-design/

Department of Education - Cordillera Administrative Region. DepEd - Cordillera Administrative Region. 2019 [cited 2022 Dec 28]. RM No. 157, s. 2019 - Result of National Achievement Test

(NAT) for School Year (SY) 2016-2017 and 2017-2018. Available from: https://www.depedcar.ph/regional-memoranda/rm-no-157-s-2019

Espinosa AA, Junio MMV, Manla MC, Palma VMS, Lucenari JLS, Punzalan AE. Analysis of achievement tests in secondary Chemistry and Biology. Int J Learn, Teach Educ Res. 2014 Apr 30;4(1):75–82. Available from https://www.ijlter.org/index.php/ijlter/article/view/11/37

Fitriyana N, Wiyarsi A, Ikhsan J, Sugiyarto KH. Android-based-game and blended learning in chemistry: effect on students' self-efficacy and achievement. Cakrawala Pendidikan. 2020 Oct 14;39(3):507–521. https://doi.org/10.21831/cp.v39i3.28335

Gabel D. Improving teaching and learning through Chemical education research: A look to the future. J Chem Educ. 1999;76(4):548–554. https://doi.org/10.1021/ed076p548

Garris R, Ahlers R, Driskell JE. Games, motivation, and learning: A Research and practice model. Simulation & Gaming. 2002 Dec;33(4):752–763. https://doi.org/10.1177/1046878102238607

Hake RR. Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. Am J Phys. 1998 Jan;66(1):64–74. https://doi.org/10.1119/1.18809

International Organization for Standardization. ISO 25010 [Internet]. Available from: https://iso25000.com/index.php/en/iso-25000-standards/iso-25010

Irwanto, Rohaeti E, Widjajanti E, Suyanta. Students' science process skill and analytical thinking ability in chemistry learning. In Yogyakarta, Indonesia; 2017 [cited 2022 Dec 22]. p. 030001. Available from: http://aip.scitation.org/doi/abs/10.1063/1.4995100

ISFET. ISFET (International Society for Educational Technology). [cited 2023 Jan 11]. ADDIE Model. Available from: https://www.isfet.org/pages/addie-model

Johnstone AH. Why is science difficult to learn? Things are seldom what they seem. J Comput Assis Learn. 1991;7(2):75–83. https://doi.org/10.1111/j.1365-2729.1991.tb00230.x

King BM, Rosopa P, Minium EW. Statistical reasoning in the behavioral sciences. Seventh edition. Hoboken, NJ: John Wiley & Sons, Inc; 2018.

Kirschner F, Paas F, Kirschner PA. Individual and group-based learning from complex cognitive tasks: Effects on retention and transfer efficiency. Comput Human Behavior. 2009 Mar;25(2):306–314. https://doi.org/10.1016/j.chb.2008.12.008

Mangiafico SS. Summary and Analysis of Extension Program Evaluation in R [Internet]. 2016. Available from: https://www.researchgate.net/publication/305436975_Summary_and_Analysis_of_Extension_Program_Evaluation_in_R

Mishra P, Pandey C, Singh U, Gupta A, Sahu C, Keshri A. Descriptive statistics and normality tests for statistical data. Ann Card Anaesth. 2019 Jan;22(1):67-72. https://doi.org/10.4103/aca.ACA_157_18

Nakthong U, Anuntasethakul T, Yutakom N. Science process skills of Grade 10 students related to cells and cell processes. 2007 Sep;13(3):383–94.

Rahmawati Y, Ramadhani SF, Afrizal A, Puspitasari M, Mardiah A. Development of students' conceptual understanding through STEAM project integration in Thermochemistry. JTK (Jurnal

Tadris Kimiya). 2021 Jun 30;6:62–74. Available from https://journal.uinsgd.ac.id/index.php/tadris-kimiya/article/view/5498

Richardson J. Why Is Chemistry So Boring? (11 Revealing Reasons) [Internet]. VeryInformed. 2022 [cited 2022 Dec 23]. Available from: https://veryinformed.com/why-is-chemistry-so-boring/

Rosal GM, Aguinaldo JCM, Reyes LDB, Casuat GHU, Balagtas RU, Del Mundo, EF. Improving the least mastered competencies of Grade 11 students in General Chemistry using electronic Strategic intervention material (E-SIM). KIMIKA 2022 Dec;33(2):59–76. https://doi.org/10.26534/kimika.v33i2.59-76

Saraswati S, Linda R, Herdini H. Development of interactive E-module Chemistry magazine based on Kvisoft Flipbook maker for Thermochemistry materials at sSecond grade Senior High School. J Sci Learn. 2019 Nov 28;3(1):1–6. Available from https://ejournal.upi.edu/index.php/jslearning/article/view/44/pdf

Seetee N, Coll RK, Boonprakob M, Dahsah C. Exploring integrated science process skills in Chemistry of high school students. Veridian E-Journal,Silpakorn University (Humanities, Social Sciences and arts). 2016 Jun;9(4):247–259. Available from https://he02.tci-thaijo.org/index.php/Veridian-E-Journal/article/download/64368/52804/149831

Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika. 1965 Dec 1;52(3–4):591–611. Available from https://www.jstor.org/stable/2333709

Shukor NA, Tasir Z, Van der Meijden H. An Examination of online learning effectiveness using data mining. Procedia - Soc Behav Sci. 2015 Jan;172:555–562. https://doi.org/10.1016/j.sbspro.2015.01.402

Sokrat H, Tamani S, Moutaabbid M, Radid M. Difficulties of students from the Faculty of Science with regard to understanding the concepts of Chemical Thermodynamics. Procedia - Soc Behav Sci. 2014 Feb 21;116:368–372. https://doi.org/10.1016/j.sbspro.2014.01.223

Solikhin F, Wijanarko A. The development of Android-based learning media (Chemdroid) on the topic Thermochemistry to improve the students' achievement. Jurnal Kimia Dan Pendidikan Kimia. 2021 Aug;6(2):138–152. Available from https://jurnal.uns.ac.id/jkpk/article/view/46849

Subhash S, Cudney EA. Gamified learning in higher education: A systematic review of the literature. Comput Hum Behav. 2018 Oct;87:192–206. https://doi.org/10.1016/j.chb.2018.05.028

Titov A. Why is Unity the best game engine? From Lead Unity Developer [Internet]. 2022 [cited 2023 Jan 11]. Available from: https://stepico.com/blog/why-is-unity-the-best-game-engine-pros-and-cons/

Wan Ahmad WF, Abdul Rahman NF. AKAMIA: Chemistry mobile game-based tutorial. In: 2014 3rd International Conference on User Science and Engineering (i-USEr) [Internet]. Shah Alam, Malaysia: IEEE; 2014 [cited 2022 Oct 24]. p. 221–6. Available from: http://ieeexplore.ieee.org/document/7002706/

Wang H, Sun CT. Game reward systems: Gaming experiences and social meanings. Proceedings of DiGRA 2011 Conference: Think Design Play. 2011. Available from https://www.researchgate.net/publication/268351726_Game_Reward_Systems_Gaming_Experiences_and_Social_Meanings

Wilcoxon F. Individual Comparisons by Ranking Methods. In: Kotz S, Johnson NL, editors. Breakthroughs in Statistics [Internet]. New York, NY: Springer New York; 1992 [cited 2023 May 27]. p. 196–202. (Springer Series in Statistics). Available from: http://link.springer.com/10.1007/978-1-4612-4380-9_16

Woolson RF. Wilcoxon Signed-Rank Test. In: D'Agostino RB, Sullivan L, Massaro J, editors. Wiley Encyclopedia of Clinical Trials [Internet]. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2008 [cited 2023 May 27]. p. eoct979. Available from: https://onlinelibrary.wiley.com/doi/10.1002/9780471462422.eoct979

Xu M, Fralick D, Zheng JZ, Wang B, Tu XM, Feng C. The differences and similarities between two-sample T-test and paired T-test. Shanghai Arch Psychiatry. 2017 Jun;29(3):184–188. Available from https://www.researchgate.net/publication/319728447_The_Differences_and_Similarities_Between_Two-Sample_T-Test_and_Paired_T-Test

York RO. Statistics for human service evaluation. First Edition. Los Angeles: SAGE Publications, Inc; 2017. 247 p.