Analyzing interaction and player experience of game based learning using feature importance based clustering

Muhammad Bahauddin Alfan [a,1], Umi Laili Yuhana [a,2,*], Darlis Herumurti [a,3]

[a] Department of Informatics, Institut Teknologi Sepuluh Nopember (ITS), Jl. Teknik Kimia, Keputih, Kec. Sukolilo, Surabaya, Jawa Timur, Indonesia
[1] 6025221028@student.its.ac.id; [2] yuhana@if.its.ac.id; [3] darlis@if.its.ac.id
* corresponding author

ARTICLE INFO

Article history
Received July 18, 2024
Revised July 24, 2024
Accepted December 8, 2024

Keywords
Digital game-based learning
FLM-GOMS
Interaction analysis
In-game experience
iGEQ
K-means clustering

ABSTRACT

This study explores the dynamics of the gaming experience and its impact on gaming interaction through digital game-based learning (DGBL). Leveraging the Fingerstroke Level Model-GOMS (FLM-GOMS) for interaction analysis and the In-Game Experience Questionnaire (iGEQ) for player experience assessment, we examine the relationship between game-play mechanics and educational outcomes. Our research incorporates a comprehensive dataset, focusing on 40 features encompassing motivation and efficiency outcomes. Through clustering, we identify distinct player groups exhibiting significant variations in interaction analysis and game experiences. We utilized the feature selection technique to identify the crucial features that differentiate groups of students who excel in implementing DGBL from those who do not. Through the Random Forest feature importance method, we have found that FLM-GOMS features and positive player in-game feedback play a pivotal role in determining the effectiveness of DGBL.

1. Introduction

Digital game-based learning (DGBL) is an educational method that utilizes games to facilitate learning and motivation [1], [2]. Students can acquire knowledge and skills through gameplay in an engaging and interactive environment [3], [4]. This approach supports active learning, enhances student engagement, and allows for the direct application of learned concepts. Digital games as a learning medium provide immediate feedback, aiding students in comprehending the consequences of their decisions and encouraging reflection on their learning [5], [6].

The implementation of effectiveness measurement using a specific framework is required to evaluate whether digital game-based learning is effective as a replacement for conventional learning processes. Various frameworks can be utilized, and in this research, referring to the conceptual framework [7]–[9] it is depicted that learning outcomes, motivation outcomes, and efficiency outcomes are crucial elements in the evaluation process of digital game-based learning effectiveness.

Based on the conceptual framework, digital game-based learning can be deemed effective if the three outcomes (learning, motivation, and efficiency) can demonstrate significant results between the conditions before and after the intervention of digital game-based learning. Digital game-based learning can take various forms, depending on the devices or technologies used. This means that the way it is implemented may vary across different devices or technologies. In particular, mobile technology greatly impacts the creation of educational games, making them more engaging and enjoyable. Consequently, there is a need for new guidelines to improve mobile game development [10]–[12], especially in the education and learning process.

https:doi.org/10.31763/businta.v8i2.772

According to a literature study conducted by Gris & Bengtson [13], out of 54 research instruments used to measure the effectiveness of DGBL in the learning process, 41 use indirect measurement. This is based on observations and Likert scale-based questionnaires before and after the intervention. However, some are also obtained from interactions between DGBL and the players' behaviors on the game features [14].

Direct interactions between gameplay and player actions, such as in-game questionnaires, cognitive scores, quiz scoring, and other instruments are classified as direct measurements. However, from the results of the literature study related to DGBL, direct measurement is an instrument that is rarely used in evaluating the performance of DGBL in terms of usability and engagement [15]–[17].

Research instruments that are characterized by indirect measurement have limitations when measuring the motivation of players in real time without disrupting the course of the game or gameplay. For example, if usability measurement is conducted after the game session ends, then a subjective perception of the players influenced by what happened during the game session will emerge [18].

The need in this research is to evaluate the overall effectiveness of the learning process both from the in-game and pre-post-game aspects. This is because perceptions related to student interactions and student behavior with the game cannot be analyzed as in instructional learning that involves direct contact with the instructor.

2. Related Work

Various studies have conducted analyses on the effectiveness of game-based learning. Some research employs data mining approaches based on the clustering of decision-making data within games, such as the number of interactions and problem-solving abilities within the game [19]–[21]. These studies yield assessments of student capabilities based on the feature selection clustering such as the frequency of successes and failures in completing specific tasks. On the other hand, some studies utilize statistics to measure the differences before and after the intervention of game-based learning [22]. These studies commonly adopt a pre-test and post-test approach, thus focusing solely on learning outcomes. Furthermore, some studies assess learning motivation through questionnaires. This approach typically explores participants' opinions, which is effectively used to gauge the depth of the subjects' learning motivation [23].

These three approaches were carried out separately by each research, so a more in-depth analysis was needed regarding the three outcomes (learning, motivation, and efficiency) [24]. According to [5], digital game-based learning can improve student learning through increased motivation and engagement. It is important to note that digital game-based learning may not be effective for students who lack motivation or interest in games. To determine the overall effectiveness of game-based learning, it is necessary to analyze how the three outcomes of learning, motivation, and efficiency are interconnected. A thorough investigation of these outcomes is crucial for evaluating the overall effectiveness of game-based learning.

2.1. Fingerstroke Level Model-GOMS (FLM-GOMS)

The Fingerstroke Level Model (FLM) is an efficient model framework that predicts the performance time of basic interaction controls (Tapping, Pointing, Dragging, and Flicking) on touch-sensitive smartphone interfaces for mobile gaming applications [25]. Essentially, FLM is developed based on a regression model, offering a different approach to estimating interaction time.

In the basic research of FLM, there are two stages of experimentation: the first stage is to establish a unit time estimation for four physical operators (Tapping, Pointing, Dragging, and Flicking) and the second stage is to validate this time estimation through game performance evaluation. Based on this research step, FLM can provide empirical evidence related to efficient interaction analysis in game development. However, there is a dynamic nature of user behavior that becomes an obstacle in FLM's basic research. This is because basic research only relies on predetermined interaction times, which may not take into account users' dynamic and evolving behaviors and preferences over time. Therefore, to analyze the effectiveness of game-based learning, other supporting factors are needed to measure effectiveness by involving user preferences and behavior.

On the other hand, GOMS (Goals, Operators, Methods, and Selection Rules) is a cognitive modeling technique that aims to analyze and predict human performance when interacting with computer systems [26]. GOMS consists of Goals (specific objectives that users want to achieve), Operators (actions required to achieve the goals, which in this case is FLM), Methods (procedures to achieve the goals), and Selection Rules (decisions made to choose among methods).

Broadly, the application of FLM-GOMS represents the development of methods to optimize the sequence of tasks or actions required to achieve a specific goal within a game and identify interaction bottlenecks. This enables game developers to create a more intuitive and responsive gaming experience.

$$Ef_{raw} = \frac{\sum opr}{T} \tag{1}$$

$$Ef_{normalized} = 1 + \frac{Ef_{raw} - \min(Ef_{raw})}{[\max(Ef_{raw}) - \min(Ef_{raw})]} * 9 \tag{2}$$

To calculate the efficiency score based on the FLM-GOMS concept (represented as Ef in Eq. 1), two primary variables are proposed, namely the number of operators performed in one sequence or game session as the numerator and the time spent in one sequence or game session as the denominator. To obtain a normal distribution of efficiency values, normalization is necessary to improve prediction accuracy and reduce the spread of points in the clustering process [27], as shown in Eq. 2.

2.2. In-game Experience Questionnaire (iGEQ)

When developing mobile game-based learning, developers must consider how users' gaming experiences affect the effectiveness of the learning process. This can also be used to gauge how users feel and act about game-based learning.

The In-Game Experience Questionnaire (iGEQ), a measurement tool aimed at evaluating a player's experience during a game session, is one tool that can be used [28]. iGEQ is applied when users have specific gaming experiences during a game session (Competence, Flow, Challenge, Tension, Positive Affect, and Negative Affect), instead of the Game Experience Questionnaire (GEQ) concept that is measured based on pre-test and post-test. However, GEQ questionnaires were still used in this work as a characteristic to support the clustering pattern and as a comparative tool between iGEQ and GEQ.

3. Method

This study involves multiple phases, including developing and implementing the Rapid Ratio game, data clustering, and evaluation of data classification. The following is the brief explanation of the clustering process using K-means which applies feature importance.

3.1. Overview of Methodology

This section describes the methodology used to cluster data based on important features derived from a Random Forest Classifier. The process involves selecting the optimal number of clusters using the Calinski-Harabasz score, performing feature selection, and iteratively refining the clusters until convergence. The pseudocode is broken down into multiple parts for clarity.

3.2. Initialize Number of Clusters (K)

To determine the optimal number of clusters (K), we use the Calinski-Harabasz score. Based on this criterion, we set K to 2.

```
Input: Dataset with 40 features (learning, motivation, and
efficiency outcomes)
Output: Clustered data with labeled clusters and important
features

Initialize K:
    Based on Calinski-Harabasz score, K = 2
```

3.3. Feature Selection Using Random Forest Classifier

We use a Random Forest Classifier to identify the most important features that contribute to the clustering process.

```
Feature Selection using Random Forest Classifier:
    a. Train Random Forest:
        random_forest = RandomForestClassifier()
        random_forest.fit(dataset.features, dataset.labels)

    b. Extract feature importances:
        feature_importances = random_forest.feature_importances_

    c. Select top N important features:
        N = 10
        important_features_indices =
select_top_N_features(feature_importances, N)
        selected_features          =          dataset.features[:,
important_features_indices]
```

3.4. Initialize Centroids

We initialize the centroids by randomly selecting K data points from the selected features:

```
Initialize centroids:
    Centroids                                                        =
random_selection_of_K_data_points(selected_features, K
```

3.5. K-means Clustering Process

The clustering process involves assigning data points to the nearest centroid and updating the centroids until they converge.

```
K-means Clustering with Feature Importance Pseudocode
Repeat until convergence:
    while not_converged:
        # Step 4a: Assign clusters
        clusters = [[] for _ in range(K)]
        for each data_point in selected_features:
            distances = []
            for each centroid in centroids:
                distance = calculate_distance(data_point,
centroid)
                distances.append(distance)
            closest_centroid =
find_min_distance_index(distances)
            clusters[closest_centroid].append(data_point)

        # Step 4b: Update centroids
        new_centroids = []
        for each cluster in clusters:
            new_centroid = calculate_mean(cluster)
            new_centroids.append(new_centroid)
```

```
        if new_centroids == centroids:
            converged = True
        else:
            centroids = new_centroids
```

3.6. Output Final Clusters and Important Features

Finally, we output the clusters, centroids, and the indices of the selected important features:

```
Output the final clusters, centroids, and selected features:
    return clusters, centroids, important_features_indices
```

This method aims to improve clustering accuracy by focusing on the most relevant features. The use of Random Forest for feature selection ensures that the clustering process emphasizes the most impactful variables, leading to more meaningful and interpretable clusters. The iterative refinement of centroids guarantees that the clustering solution is optimal and stable.

A proposed framework for evaluating the effectiveness of digital game-based learning is shown in Fig. 1.

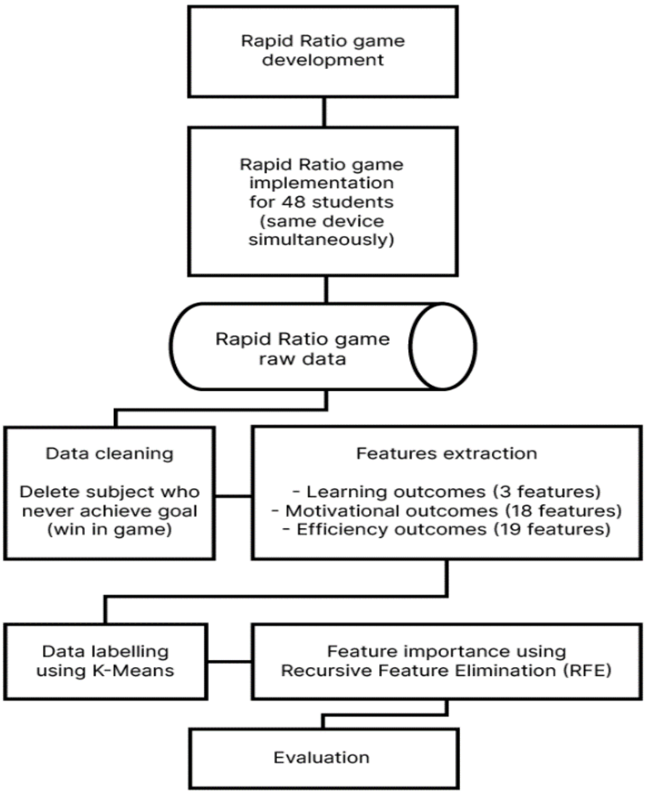

Fig. 1. The proposed framework to assess the effectiveness of digital game-based learning

3.7. Overview of Rapid Ratio Game

Based on one of the fundamental mathematical skills"explaining the ratio of two quantities (same and different units) with a focus on scale and proportion factors"these numeracy materials were chosen to develop digital game-based learning in this study. This mobile game-based learning, also known as "Rapid Ratio", educates players on how to apply the concepts of time and speed to reach the greatest score while facing challenges throughout the particular gameplay.

The application of operators from the previously explained Fingerstroke Level Model-GOMS is an essential component of the game development process. Players in the game have limited energy to move around using the Dragging operator (D). To get a higher score, players should use this operator sparingly and plan their movements carefully to overcome obstacles. Additionally, players can use the Tapping (T) operator to shoot enemies and earn points, the Pointing (P) operator to pick up or place items, and the Flicking (F) operator to jump over obstacles. The player will replay the game session and go back to its starting position if its energy runs out, get shot by an enemy or failed to avoid an obstacle. In the game "Rapid Ratio," two distinct gameplay mechanics are employed: object placement and ball shooter. Object placement as shown in Fig. 2, within the context of this game, is defined as the player required to locate blue cubic objects in proportions that match the predetermined ratio to complete the game.

Fig. 2. Object placement gameplay in Rapid Ratio

For instance, as illustrated in Fig. 2, the player is presented with a ratio of 2:1 area along with a “remaining” area for surplus objects that could potentially alter the proportionality of the objects' ratio. When dealing with a 2:1 ratio and having 10 blue cubes available, there are three possible solutions one can use. The higher the players minimize the “remaining” area, the higher their knowledge score. Ball shooter gameplay in Rapid Ratio as show in Fig. 3.

- The first solution involves placing 6 blue cubes in the "2" area, 3 blue cubes in the "1" area, and 1 blue cube in the "remaining" area. This is the highest-scoring solution, so the player deserves 100% additional energy.
- The second solution requires placing 4 blue cubes in the "2" area, 2 blue cubes in the "1" area, and 4 blue cubes in the "remaining" area. This is the second-highest-scoring solution, so the player will get 70% additional energy.
- The third solution involves placing 2 blue cubes in the "2" area, 1 blue cube in the "1" area, and 7 blue cubes in the "remaining" area. This is the lowest-scoring solution, so the player will get 40% additional energy.

Fig. 3. Ball shooter gameplay in Rapid Ratio

Furthermore, the ball shooter gameplay involves the player facing enemies capable of shooting at the player. If the player gets hit, the game session restarts from the initial position. Consequently, the player must eliminate enemies by shooting them with projectiles corresponding to the fractional values the enemies represent.

In this scenario (Fig. 3), the player is confronted with an enemy who has a fractional life value of 12/25, which is equivalent to 0.48 in decimal form. To win the game, the player has to shoot a bullet that reduces the enemy's life to zero. However, if the player chooses the wrong bullet, the enemy's life may end up being negative, and the enemy could still attack the player. So the player must conserve energy and maximize scores by strategic bullet selection.

The game "Rapid Ratio" has been developed using the Unity engine. The development process was divided into three stages. In the first stage, game developers discussed the gameplay with mathematics experts to determine if it was worth developing and testing on participants. In the second stage, a game prototype was developed and tested on experts as well as a sample of students to assess how the game mechanics worked according to users. The final stage involved debugging and fixing mechanical or gameplay errors based on suggestions received during the second stage.

3.8. Data Collection

The experiment to test the use of mobile game-based learning was carried out in classrooms using mobile devices with the same specifications, namely the SPC Tablet L80 LITE 4G 8 inch. The experiment was carried out by 48 students aged between 12 and 14. The data collection process was divided into three phases.

In the first phase, players received an explanation of how to play from tutors and teachers, including a tutorial session conducted together. During the second phase, participants were asked to play for 30 minutes while accompanied by teachers who would only help if technical problems occurred, but they did not help in answering questions in the game. The third phase involved collecting data from player logs and distributing questionnaires that are embedded in the "Rapid Ratio" game. Data from these logs will be further analyzed to understand how the game interaction functions from the player's perspective.

3.9. Feature Extraction

The effectiveness of digital game-based learning outcomes is divided into 40 features with the following details.

- Learning outcome: 2 features. These two features come from assessing knowledge of playing games and mathematics regarding ratios and proportions before and after DGBL implementation.
- Motivational outcome: 18 features. Based on six different Game Experience Questionnaires, these 12 features were taken before and after the intervention. The rest 6 features/questionnaires were also taken during the direct DGBL implementation process using iGEQ.
- Efficiency outcome: 20 features. One feature comes from an efficiency score that is calculated using FLM-GOMS. The other feature comes from the game competency score that is not formulated using FLM-GOMS. The 16 features are derived by calculating the time average of each operator per player, regardless of whether it is from the lost state or win state. The remaining two features are defined by adding up the winning state achieved and the lost state.

These 40 outcomes are features for finding the relationship between students' learning outcomes, motivation outcomes, and efficiency outcomes in completing digital game-based learning.

3.10. Data Labelling using K-means

To better understand the raw data that would be extracted from the "Rapid Ratio" game, it is necessary to label 40 features and analyze any patterns that emerge for each subject. This will help determine if a digital game-based approach is effective across all outcomes or only in certain outcomes. However, since each outcome (or feature) has a different value or scale, clustering is necessary to identify which clusters are labeled as effective groups and which are not [29]. After clustering, it is important to analyze the centroid values to identify the features that have a high influence in creating the clusters. By doing this, a better overview could be obtained regarding whether a subject is effectively implementing game-based learning or not.

Thus, in this case, using the K-means clustering algorithm is advantageous for its computational efficiency, particularly with large datasets and features. Its simplicity and effectiveness are well-regarded in various applications, including data mining and predictive modeling [30].

Vol. 8, No. 2, December 2024, pp. 295-308

3.11. Feature Importance Selection

This study aimed to evaluate and compare the data labeling quality of a given dataset by employing two distinct feature selection methods: Random Forest Classifier (RFC) and Support Vector Regression (SVR) with a linear kernel. The Calinski-Harabasz score was utilized as the primary metric for assessing clustering quality. The methodology can be divided into several key stages:

- Data Preparation: The dataset was initially loaded and cleaned by removing specific clustering-related columns to ensure an unbiased comparison. Missing values were imputed with the median of each column to maintain data integrity.
- Feature Selection: Two feature selection methods were employed to identify the most significant features for clustering analysis. The Random Forest Classifier (RFC) and Support Vector Regression (SVR) with a linear kernel were utilized, each set to select the top 10 features. The RFC method leverages an ensemble of decision trees to assess feature importance, while SVR identifies features that contribute significantly to the prediction of cluster labels.
- Clustering Quality Assessment: With the selected features from both methods, the clustering quality was assessed using the Calinski-Harabasz score, a metric that evaluates the ratio of between-cluster dispersion to within-cluster dispersion. Higher scores indicate better clustering quality, with clusters being more distinct and compact.
- Statistical Analysis: The study used statistical analysis to compare the Calinski-Harabasz scores resulting from the two feature selection methods, aiming to determine which method and which features led to better clustering outcomes.

Through this methodology, the study seeks to provide insights into the effectiveness of feature selection methods in improving clustering quality.

4. Results and Discussion

In this section, we will detail the experimental results and metrics adopted to measure our proposed framework's performance.

4.1. Data Analysis

Before the launch of the "Rapid Ratio" game, all participants were split into two large groups. The first group, consisting of 19 students, was the "Low Performing Group". These students had relatively low initial knowledge of mathematics and were less skilled at playing the game. The second group, called the "High Performing Group", comprised 17 students who had higher initial knowledge of mathematics and were more skilled at playing the game. Based on the initial data, we can observe that the "Low Performing Group" had a lower efficiency score (based on interaction analysis) and motivation score (based on in-game experience questionnaire) as compared to the "High Performing Group". Efficiency (FLM-GOMS) and Motivation (iGEQ) score difference between two groups as show in Fig. 4.

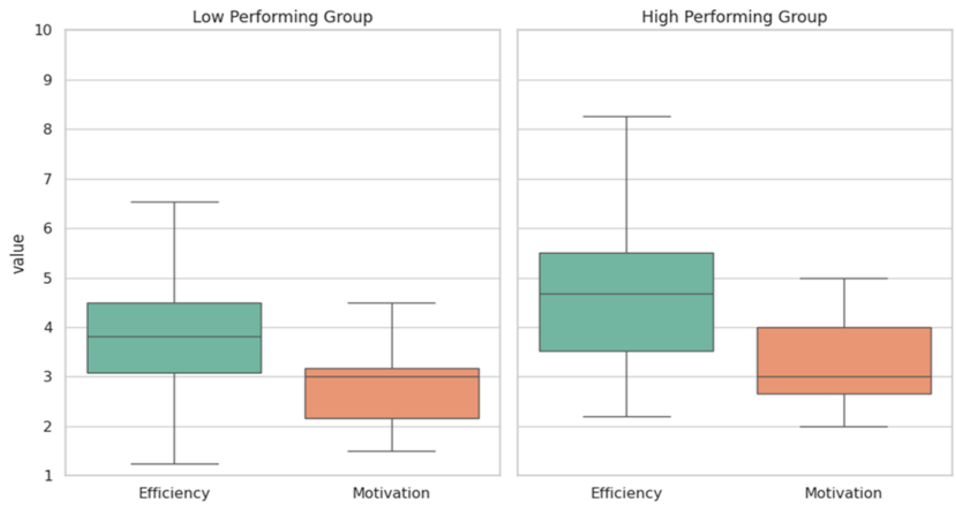

Fig. 4. Efficiency (FLM-GOMS) and Motivation (iGEQ) score difference between two groups

According to Fig. 4, the efficiency score calculated using FLM-GOMS is more effective than the motivation score in differentiating between student groups. The motivation score does not have a different median between the Low Performing Group and the High Performing Group. However, this preliminary analysis alone is not sufficient to prove that FLM-GOMS can accurately group efficient and inefficient participants in digital game-based learning. Further analysis is required to determine the differences.

4.2. Data Clustering Results

To label data using clustering, it is important to determine the optimal number of clusters. The Calinski-Harabasz score is a measure of cluster validity that indicates the validity of the clusters. Higher scores indicate better-defined clusters. Based on the scores mentioned in Table 1, dividing the dataset into 2 clusters would result in the most distinct and well-separated clusters. The Calinski-Harabasz scores for a range of clusters from 2 to 10 are provided in Table 1 for reference.

Table.1 Calinski-Harabasz Scores of 2-10 Clusters

Feature	Value
2	**5.80**
3	4.82
4	4.13
5	4.01
6	3.27
7	3.38
8	3.12
9	3.07
10	2.94

Component Using K-means, the data clustering process yielded two distinct clusters which are labeled as Cluster 0 and Cluster 1. Cluster 0 comprises 23 students, of which 8 are male and 15 are female students. This cluster consists of players who have relatively lower performance and hold lesser positive perceptions about their playing experience. On the other hand, Cluster 1 is composed of 13 students, of which 12 are male and 1 is female. This cluster primarily consists of players who have better performance and possess stronger positive perceptions about their competence and overall playing experience.

To determine the nature of the two clusters, we can analyze the differences in centroid values between them. By conducting centroid analysis, we can identify the main factors that influenced cluster formation. This is based on the difference in values between centroids for each feature presented in Table 2.

Table.2 Value Differences Between Centroids for each Feature

Feature	Value
Efficiency score	1.29
Competence GEQ post-game	1.17
Flow GEQ post-game	1.17
The time average of Pointing (P) on lost state	0.94
The time average of Pointing (P) on win state	0.82
Competence iGEQ	0.67
Flow iGEQ	0.65
Positive Affect GEQ post-game	0.64
Tension GEQ post-game	0.52

The following features were observed in two different clusters of players:

- Efficiency score: there was a significant difference in playing efficiency between the two clusters, with cluster 1 showing higher efficiency than cluster 0.

- Competence GEQ post-game and Flow GEQ post-game: there was a large difference in perceived competence and experience of flow after playing. Cluster 1 players had a more positive gaming experience and felt more competent.
- The time average of Pointing (P) on the win state and The time average of Pointing (P) on the lost state: there were differences in the average duration of losing and winning on P-type interactions between the two clusters, highlighting how performance in these specific interactions differentiates clusters.
- Competence iGEQ and Flow iGEQ: perceptions of general competence and flow experiences also differed significantly. This indicates variations in how players perceive their abilities and experiences during play.

The above differences suggest that certain variables, such as efficiency outcomes and emotional responses after playing, play a key role in grouping players. Cluster 1 tends to include players with better performance and stronger positive perceptions of their competence and playing experience. On the other hand, Cluster 0 includes players with relatively lower performance and perhaps less positive perceptions of their playing experience.

4.3. Feature Selection Results

In the exploration of feature selection methods for enhancing clustering performance, our analysis employed three distinct approaches: Recursive Feature Elimination (RFE) with a Random Forest Classifier estimator, RFE with a Support Vector Regression (SVR) estimator, and feature importance based on Random Forest Classifier. The Calinski-Harabasz score, a metric indicative of cluster quality through the evaluation of cluster density and separation, served as our evaluation criterion. The Calinski-Harabasz score for the RFE-RFC approach was 8.1886, while the RFE-SVR approach produced a score of 4.4734. The best score was achieved by utilizing feature importance based on RFC, which yielded a score of 8.57.

Based on Fig. 5, this finding highlights the effectiveness of leveraging Random Forest-based feature selection in achieving more coherent and distinct clustering outcomes, thereby providing valuable insight into the optimization of clustering techniques in data analysis.

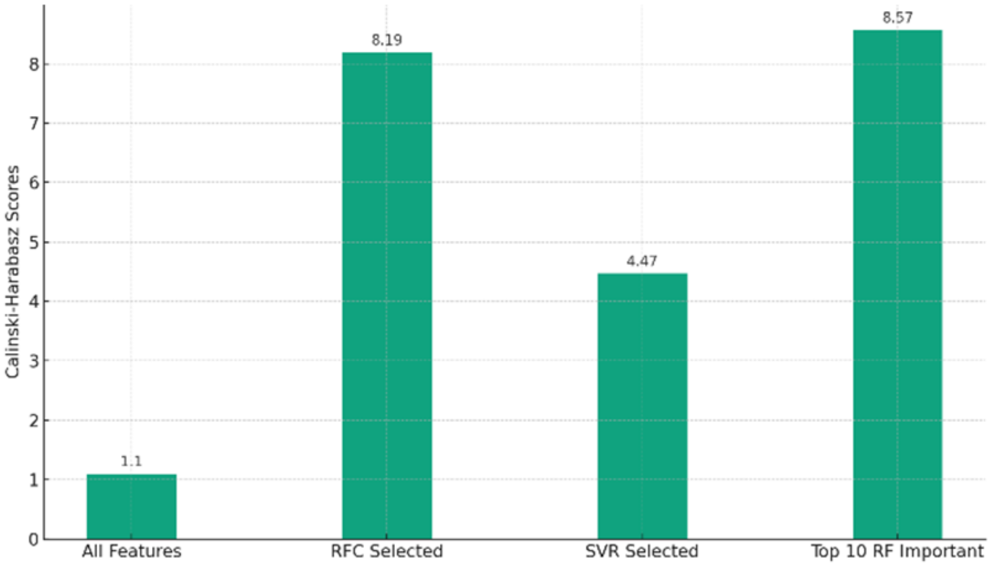

Fig. 5. Calinski-Harabasz Score by non-selection method and feature selection method

This notable discrepancy underscores the superior clustering performance achieved through the features selected by the Random Forest Classifier. Specifically, the higher Calinski-Harabasz score associated with the Random Forest Classifier-selected features suggests that these clusters are significantly denser and more distinctly separated compared to those derived from SVR-selected features. This finding highlights the effectiveness of leveraging Random Forest-based feature selection in achieving more coherent and distinct clustering outcomes, thereby providing valuable insight into the optimization of clustering techniques in data analysis.

It's interesting to note that features included in the efficiency outcome (highlighted in Table 3) are chosen through the utilization of feature importance techniques based on the Random Forest Classifier. Thus, this supports the theory that efficiency outcomes play a significant role in grouping students who implement DGBL successfully.

Table.3 Top 10 Feature Selection using 3 Different Methods

RFE estimator = RFC	RFE estimator = SVR	Feature Importance (Random Forest Classifier)
Tension GEQ post-game	Game competency score	The time average of Dragging Up (DU) on win state
Challenge iGEQ	Competence iGEQ	The time average of Dragging Up (DU) on lost state
Negative Affect iGEQ	Competence GEQ post-game	The time average of Dragging Left (DL) on win state
The time average of Dragging Up (DU) on lost state	Flow GEQ post-game	The time average of Dragging Down (DD) on win state
The time average of Idle (DU) on lost state	The time average of Dragging Up (DU) on lost state	The time average of Pointing (P) on lost state
The time average of Pointing (P) on lost state	The time average of Pointing (P) on lost state	Flow GEQ post-game
The time average of Dragging Down (DD) on win state	The time average of Dragging Down (DD) on win state	The time average of Dragging Left (DL) on lost state
The time average of Dragging Left (DL) on win state	The time average of Dragging Left (DL) on win state	Competence GEQ post-game
The time average of Dragging Right (DR) on win state	The time average of Dragging Up (DU) on win state	The time average of Pointing (P) on win state
The time average of Dragging Up (DU) on win state	The time average of Pointing (P) on win state	Efficiency score

4.4. Evaluation

In In the evaluation section of this research, we rigorously assessed the performance of two machine learning models, Random Forest and Support Vector Machines (SVM), across various feature selection methods, including all features, features selected by Random Forest Classifier (RFC), Support Vector Regression (SVR) selected features, and the top 10 features identified by importance through Random Forest. Calinski-Harabasz Score by non-selection method and feature selection method as show in Fig. 6.

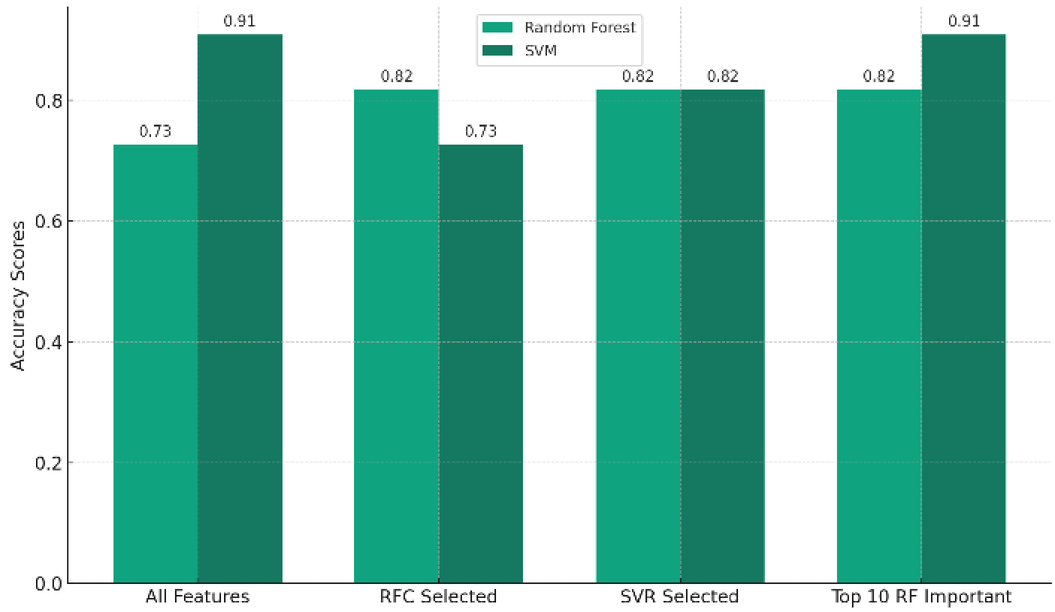

Fig. 6. Calinski-Harabasz Score by non-selection method and feature selection method

The models were evaluated based on accuracy metrics derived from the classification reports as shown in Fig. 6. The SVM model demonstrated remarkable adaptability and superior performance, particularly when leveraging all features and the top 10 RF important features, achieving an accuracy of 90.91% in both scenarios. This suggests SVM's robustness in handling both high-dimensional data and optimized feature subsets. Conversely, the Random Forest model exhibited a notable enhancement in accuracy with feature selection, underscoring the importance of feature selection in improving model efficacy. The consistency in Random Forest's performance improvement with RFC, SVR, and its top 10 important features indicates a direct correlation between targeted feature selection and model accuracy. This comprehensive evaluation elucidates the significant impact of feature selection on model performance, offering valuable insights for optimizing machine learning workflows in predictive modeling and classification tasks.

4.5. Limitations

Despite the significant findings and contributions of this study on digital game-based learning (DGBL), several limitations should be acknowledged:

- Sample Size and Demographics: The study was conducted with a relatively small sample size of 48 students aged between 12 and 14. This limited sample may not fully represent the broader population of students who could benefit from DGBL. Future research should consider involving a larger and more diverse sample to enhance the generalizability of the findings.
- Short-Term Evaluation: The data collection process was limited to a single 30-minute game session. This short-term evaluation might not capture the long-term effects of DGBL on learning outcomes, motivation, and efficiency. Longitudinal studies are needed to assess the sustained impact of DGBL over extended periods.
- Device and Environment Constraints: The study was conducted using SPC Tablet L80 LITE 4G devices within a controlled classroom environment. The findings may vary if different devices or environments are used. It is important to test the DGBL approach in various settings and with different technological devices to determine its broader applicability.
- Game-Specific Findings: The results drawn from the "Rapid Ratio" game might be specific to this particular game and its design. Generalizing these findings to other educational games requires caution, as different games may have unique dynamics and impacts on learning.

By acknowledging these limitations, future research can be better directed to address these issues, thereby enhancing the robustness and applicability of findings in the field of digital game-based learning.

5. Conclusion

In this study, we explored the effectiveness of digital game-based learning (DGBL) by examining various educational outcomes. We utilized the Fingerstroke Level Model-GOMS (FLM-GOMS) for interaction analysis and the In-Game Experience Questionnaire (iGEQ) to evaluate player experiences. Our research yielded several key insights into the learning process through gameplay:

- Enhanced Learning Outcomes: Our research highlights the potential of digital games to significantly improve learning outcomes when designed to align with specific educational objectives. Games that integrate educational goals can enhance students' efficiency and motivation, leading to a more engaging learning experience.
- Distinct Clusters: By analyzing our data, we identified two distinct clusters:
  - Above-Average Effectiveness: This cluster includes students who demonstrated higher efficiency outcomes in game-based learning.
  - Below-Average Effectiveness: This cluster consists of students with lower efficiency outcomes in game-based learning.

Efficiency outcomes were the primary factors influencing the formation of these clusters.

- Feature Selection: Our use of feature selection techniques, including short touchscreen interaction time and positive sentiment features from in-game experience questionnaires, improved the accuracy of clustering DGBL effectiveness. These features were crucial in distinguishing between students who excel in DGBL and those who do not.

References

[1] J. Schuldt and H. Niegemann, "Instructional Design for Digital Game-Based Learning," Springer, Cham, 2021, pp. 299–319, doi: 10.1007/978-3-030-75142-5_13.

[2] R. Al-Azawi, F. Al-Faliti, and M. Al-Blushi, "Educational Gamification Vs. Game Based Learning: Comparative Study," *Int. J. Innov. Manag. Technol.*, pp. 131–136, 2016, doi: 10.18178/ijimt.2016.7.4.659.

[3] T. E. Coleman and A. G. Money, "Student-centred digital game–based learning: a conceptual framework and survey of the state of the art," *High. Educ.*, vol. 79, no. 3, pp. 415–457, Mar. 2020, doi: 10.1007/s10734-019-00417-0.

[4] K. Sanodariya *et al.*, "Game-based Learning for Engineering Education: Supplementing Basic Electronics Instruction with Educational Games," in *2022 10th International Conference on Information and Education Technology (ICIET)*, Apr. 2022, pp. 140–144, doi: 10.1109/ICIET55102.2022.9779011.

[5] W. A. Sadera, Q. Li, L. Song, and L. Liu, "Digital Game-Based Learning," *Comput. Sch.*, vol. 31, no. 1–2, pp. 1–1, Apr. 2014, doi: 10.1080/07380569.2014.879801.

[6] M. H. Hussein, S. H. Ow, L. S. Cheong, M.-K. Thong, and N. Ale Ebrahim, "Effects of Digital Game-Based Learning on Elementary Science Learning: A Systematic Review," *IEEE Access*, vol. 7, pp. 62465–62478, 2019, doi: 10.1109/ACCESS.2019.2916324.

[7] A. All, E. P. Nuñez Castellar, and J. Van Looy, "Towards a conceptual framework for assessing the effectiveness of digital game-based learning," *Comput. Educ.*, vol. 88, pp. 29–37, Oct. 2015, doi: 10.1016/j.compedu.2015.04.012.

[8] A. All, E. P. Nunez Castellar, and J. Van Looy, "Measuring Effectiveness in Digital Game-Based Learning: A Methodological Review.," *Int. J. Serious Games*, vol. 1, no. 2, Jun. 2014, doi: 10.17083/ijsg.v1i2.18.

[9] A. J. A. Seyderhelm, K. L. Blackmore, and K. Nesbitt, "Towards Cognitive Adaptive Serious Games: A Conceptual Framework," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11863 LNCS, Springer, 2019, pp. 331–338, doi: 10.1007/978-3-030-34644-7_27.

[10] A. Tlili, F. Essalmi, and M. Jemni, "Design of Educational Games: The Evolution from Computers to Mobile Devices," in *Lecture Notes in Electrical Engineering*, vol. 381, Springer, Cham, 2016, pp. 151–159, doi: 10.1007/978-3-319-30298-0_16.

[11] F. Gao, L. Li, and Y. Sun, "A systematic review of mobile game-based learning in STEM education," *Educ. Technol. Res. Dev.*, vol. 68, no. 4, pp. 1791–1827, Aug. 2020, doi: 10.1007/s11423-020-09787-0.

[12] J. Huizenga, W. Admiraal, G. ten Dam, and J. Voogt, "Mobile game-based learning in secondary education: Students' immersion, game activities, team performance and learning outcomes," *Comput. Human Behav.*, vol. 99, pp. 137–143, Oct. 2019, doi: 10.1016/j.chb.2019.05.020.

[13] G. Gris and C. Bengtson, "Assessment Measures in Game-based Learning Research," *Int. J. Serious Games*, vol. 8, no. 1, pp. 3–26, Mar. 2021, doi: 10.17083/ijsg.v8i1.383.

[14] J. Kang, M. Liu, and W. Qu, "Using gameplay data to examine learning behavior patterns in a serious game," *Comput. Human Behav.*, vol. 72, pp. 757–770, Jul. 2017, doi: 10.1016/j.chb.2016.09.062.

[15] I. Ghergulescu and C. H. Muntean, "Measurement and Analysis of Learner's Motivation in Game-Based E-Learning," in *Assessment in Game-Based Learning*, New York, NY: Springer New York, 2012, pp. 355–378, doi: 10.1007/978-1-4614-3546-4_18.

[16] Z. Zainuddin, M. Shujahat, H. Haruna, and S. K. W. Chu, "The role of gamified e-quizzes on student learning and engagement: An interactive gamification solution for a formative assessment system," *Comput. Educ.*, vol. 145, p. 103729, Feb. 2020, doi: 10.1016/j.compedu.2019.103729.

[17] A. Khan, F. H. Ahmad, and M. M. Malik, "Use of digital game based learning and gamification in secondary school science: The effect on student engagement, learning and gender difference," *Educ. Inf. Technol.*, vol. 22, no. 6, pp. 2767–2804, Nov. 2017, doi: 10.1007/s10639-017-9622-1.

[18] C. Luce and J. P. Kirnan, "Using Indirect vs. Direct Measures in the Summative Assessment of Student Learning in Higher Education," *J. Scholarsh. Teach. Learn.*, vol. 16, no. 4, pp. 75–91, Aug. 2016, doi: 10.14434/josotl.v16i4.19371.

[19] J. L. Sabourin and J. C. Lester, "Affect and Engagement in Game-BasedLearning Environments," *IEEE Trans. Affect. Comput.*, vol. 5, no. 1, pp. 45–56, Jan. 2014, doi: 10.1109/T-AFFC.2013.27.

[20] A. Slimani, F. Elouaai, L. Elaachak, O. Bakkali Yedri, M. Bouhorma, and M. Sbert, "Learning Analytics Through Serious Games: Data Mining Algorithms for Performance Measurement and Improvement Purposes," *Int. J. Emerg. Technol. Learn.*, vol. 13, no. 01, p. 46, Jan. 2018, doi: 10.3991/ijet.v13i01.7518.

[21] H. Jeon, H. He, A. Wang, and S. Spooner, "Modeling Student Performance in Game-Based Learning Environments," *arXiv*, pp. 1–11, Sep. 2023. [Online]. Available at: https://arxiv.org/abs/2309.13429v1.

[22] C. Wardoyo, Y. Dwi Satrio, and D. Ma'ruf, "Effectiveness of Game-Based Learning – Learning in Modern Education," *KnE Soc. Sci.*, vol. 4, no. 7, pp. 81-87–81–87, Apr. 2020, doi: 10.18502/kss.v4i7.6844.

[23] M. Nadeem, M. Oroszlanyova, and W. Farag, "Effect of Digital Game-Based Learning on Student Engagement and Motivation," *Computers*, vol. 12, no. 9, p. 177, Sep. 2023, doi: 10.3390/computers12090177.

[24] A. All, E. P. Nuñez Castellar, and J. Van Looy, "Assessing the effectiveness of digital game-based learning: Best practices," *Comput. Educ.*, vol. 92–93, pp. 90–103, Jan. 2016, doi: 10.1016/j.compedu.2015.10.007.

[25] A. Lee, K. Song, H. B. Ryu, J. Kim, and G. Kwon, "Fingerstroke time estimates for touchscreen-based mobile gaming interaction," *Hum. Mov. Sci.*, vol. 44, pp. 211–224, Dec. 2015, doi: 10.1016/j.humov.2015.09.003.

[26] K. Wang, K. Li, J. Gao, B. Liu, Z. Fang, and W. Ke, "A Quantitative Evaluation Method of Software Usability Based on Improved GOMS Model," in *2021 IEEE 21st International Conference on Software Quality, Reliability and Security Companion (QRS-C)*, Dec. 2021, pp. 691–697, doi: 10.1109/QRS-C55045.2021.00104.

[27] P. Sarkar and P. J. Bickel, "Role of normalization in spectral clustering for stochastic blockmodels," *Ann. Stat.*, vol. 43, no. 3, pp. 962–990, Jun. 2015, doi: 10.1214/14-AOS1285.

[28] K. Poels, Y. A. W. de Kort, and W. A. IJsselsteijn, "D3.3 : Game Experience Questionnaire: development of a self-report measure to assess the psychological impact of digital games." Technische Universiteit Eindhoven, p. 46, 2007. [Online]. Available at: https://research.tue.nl/en/publications/d33-game-experience-questionnaire-development-of-a-self-report-me.

[29] C. Silvestre, M. Cardoso, and M. Figueiredo, "Clustering and Selecting Categorical Features," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8154 LNAI, Springer, Berlin, Heidelberg, 2013, pp. 331–342, doi: 10.1007/978-3-642-40669-0_29.

[30] P. Pandey and I. Singh, "Comparison between Standard K-Mean Clustering and Improved K-Mean Clustering," *Int. J. Comput. Appl.*, vol. 146, no. 13, pp. 39–42, Jul. 2016, doi: 10.5120/ijca2016910868.