Optimizing Social and Emotional Learning through Modified Gale-Shapley Algorithm for Collaborative and Competitive Education

Luiz Carlos Pinheiro Junior[1,2,*], Everton Gomede[1,3], and Leonardo de Souza Mendes[1]
[1]DECOM, Faculty of Electrical and Computer Engineering (FEEC), State University of Campinas (UNICAMP), Campinas, SP, Brazil
[2]Federal Institute of Paraná (IFPR), Telêmaco Borba, PR, Brazil
[3]University of British Columbia (UBC), Vancouver, BC, Canada
Email: luiz.pinheiro@ifpr.edu.br (L.C.P.J.); everton.gomede@ubc.ca (E.G.); lmendes@unicamp.br (L.D.S.M.)
*Corresponding author
Manuscript received August 13, 2024; revised September 4, 2024; accepted September 18, 2024; published November 12, 2024

***Abstract*—Collaborative and competitive learning is essential in educational development, enhancing students' social and emotional competencies critical for academic and personal success. This study explores the integration of the Gale-Shapley algorithm, initially designed for the stable marriage problem, to optimize student pairings in collaborative and competitive learning environments. The objective is to maximize the effectiveness of Social and Emotional Learning (SEL) interventions by fostering productive social interactions and essential skill development. We propose a modified version of the Gale-Shapley algorithm to handle a single list of students, utilizing compatibility scores based on inverse Euclidean distance, Jaccard similarity, and cosine similarity. Our methodology includes generating synthetic datasets to simulate various educational contexts and evaluate the algorithm's performance. Results demonstrate the algorithm's efficiency in forming stable and effective pairs, significantly enhancing the learning experience. This innovative approach aligns with SEL guidelines and contemporary educational requirements, offering a robust, personalized, and dynamic learning framework. Future research should focus on empirical validations in diverse educational settings to confirm the algorithm's effectiveness and scalability.**

***Keywords*—collaborative learning, competitive learning, Gale-Shapley algorithm, social and emotional learning, student pairing, educational optimization**

I. INTRODUCTION

Collaborative and competitive learning is pivotal in modern education, fostering social and emotional competencies critical to academic achievement and personal growth. Social and Emotional Learning (SEL) interventions in schools have been extensively studied, with evidence supporting their positive impact on students' overall development, including academic performance, emotional regulation, and interpersonal skills [1, 2]. These interventions enhance students' resilience, empathy, and conflict-resolution abilities, essential for long-term success [3, 4]. Incorporating collaborative and adaptive learning methods has increased student engagement by creating dynamic, interactive environments that cater to individual needs [5]. Collaborative learning improves academic outcomes and fosters critical thinking, effective communication, and teamwork skills [6]. On the other hand, competitive learning scenarios boost student motivation and engagement, often driving higher levels of performance and personal achievement [7]. Both approaches offer distinct benefits, such as improved social skills and increased individual motivation [8]. Moreover, integrating technology in adaptive learning environments further enhances educational experiences by offering personalized, interactive opportunities that deepen learning [5]. The growing body of research also underscores the broader benefits of SEL programs, which include reductions in behavioral issues and improvements in students' mental health. These findings are corroborated by longitudinal studies indicating that SEL interventions lead to higher academic success and better socio-emotional adaptation in later life [3, 4]. Such outcomes emphasize the need for educational systems to prioritize interpersonal skill development alongside traditional academic instruction, focusing on preparing students for the complexities of the 21st-century workforce [1, 6].

As we navigate this evolving educational landscape, integrating collaborative and competitive learning within curricula emerges as a strategy to equip students with essential skills such as teamwork, leadership, and adaptability [9]. Through group projects and peer assessments, students develop a sense of responsibility. They are better prepared to work in diverse, interdisciplinary teams—skills highly valued in fast-changing sectors like technology [10]. Central to these activities is forming student pairs or groups, a process that can significantly influence learning outcomes. One established method for pair formation is the Gale-Shapley algorithm, initially designed to solve the Stable Marriage Problem (SMP) [11]. This algorithm has been successfully adapted for various applications, ranging from kidney exchange programs [12] to e-commerce systems [13] and school choice mechanisms [14], as well as allocating students to industry placements [15]. However, when applied to educational contexts, particularly in SEL, the Gale-Shapley algorithm presents certain limitations. For example, it assumes equal group sizes and fixed preferences—assumptions that may not hold in dynamic classroom environments [16, 17]. Given these challenges, this paper proposes modifying the Gale-Shapley algorithm, explicitly tailored for optimizing student pairings in collaborative and competitive learning activities. Our goal is to enhance the effectiveness of SEL interventions by promoting meaningful social interactions and fostering the development of vital interpersonal skills. This modification draws upon recent studies that emphasize the need for flexible, context-sensitive approaches to student grouping [16, 17]. The remainder of this paper is structured as follows: Section II outlines the theoretical foundation for our proposed modification of the Gale-Shapley algorithm. Section III presents the

doi: 10.18178/ijiet.2024.14.11.2179

methodology. Section IV analyzes the algorithm's effectiveness and discusses this approach's practical implications. Section V conclusions and suggests future research directions.

II. Literature Review

Integrating collaborative and competitive learning has been crucial in advancing educational practices, with a notable impact on student's social and emotional competencies. These competencies are of paramount importance for students' academic and personal success. Studies have indicated that SEL interventions within school settings yield favorable outcomes in students' comprehensive development, thereby establishing the potential of collaborative and adaptive methodologies as a promising approach in primary education. SEL is a crucial aspect of contemporary education, emphasizing the development of self-awareness, self-management, social awareness, relationship skills, and responsible decision-making [3]. SEL programs can significantly enhance students' social behaviors, reduce emotional stress, and improve academic performance [2]. Educators can facilitate holistic student development by integrating SEL into collaborative and competitive learning environments.

Collaborative learning involves students working together to achieve shared academic goals, promoting interdependence and mutual support [18, 19]. This approach encourages the development of critical thinking and communication skills and increases student engagement and achievement [19]. In contrast, competitive learning uses rivalry to motivate students to excel individually, promoting a drive for personal fulfillment and resilience [18]. Both methodologies have distinct advantages, and their combined application can cater to different preferences and learning objectives [18–20].

The Gale-Shapley algorithm, developed by David Gale and Lloyd Shapley in 1962, addresses the stable marriage problem, which seeks to identify a stable pairing between two sets of elements of equal size, given an order of preference for each component [11]. Consider a pairing stable if no two elements prefer each other over their current partners. Researchers have successfully adapted the Gale-Shapley algorithm for several real-world applications, including university admissions [21], medical resident placements [22], and kidney exchange programs [12]. Furthermore, the work by Fenoaltea *et al.* [23] provides an interdisciplinary review of the stable marriage problem, emphasizing its applications in various scientific fields, including statistical physics, economics, and game theory. The authors employ tools from statistical mechanics to investigate both the search for stable solutions that guarantee the system's stability and the search for global optimal solutions that maximize the total happiness of the agents involved. Also, this review serves to reinforce the importance of SMP in the modeling of two-sided markets and its relevance to the understanding of complex systems.

In the educational domain, algorithmic pairing has demonstrated significant benefits. For example, Girard *et al.* [16] developed a collaborative e-learning platform for privacy education that leverages a personalized stable pairing algorithm. This system enhances interaction by pairing students based on academic and behavioral profiles, fostering a more efficient and satisfying learning environment. Similarly, Yusri, Abusitta, and Aïmeur [17] introduced the "Teens-Online" platform, which employs game theory and stable pairing algorithms to create a collaborative privacy education space for teenagers. Considering academic and behavioral factors, these platforms ensure optimal student pairings, leading to improved educational outcomes and enhanced social skills development.

Traditional implementations of the Gale-Shapley algorithm require two distinct lists of participants, which can introduce challenges in classroom pairing, where all individuals belong to the same group [21, 22]. Roth [22] highlighted the limitations of this approach, particularly in cases where preferences are poorly defined, leading to imbalances and suboptimal pairings.

A. Proposed Modifications to the Gale-Shapley Algorithm

This study modifies the Gale-Shapley algorithm that adapts it for a single list of participants, a more realistic reflection of classroom dynamics where students are all part of the same cohort. The proposed single-list approach eliminates the need to divide individuals into two distinct sets, simplifying the pairing process and addressing the imbalances in traditional methods.

This one-list framework forms pairs based on objective compatibility metrics, such as the Jaccard coefficient, inverse Euclidean distance, and Cosine similarity. These metrics replace subjective preference rankings, providing a more precise compatibility assessment. The result is a more stable and efficient pairing process that promotes effective interaction and enhances educational outcomes.

B. Comparison with Existing Methods

While studies like those by Yusri *et al.* [16] and Girard *et al.* [17] utilize pairing algorithms grounded in game theory to facilitate stable matches, they rely on two preference lists. The approach leveraging a single-list model and objective compatibility metrics proposed in this study offers a more practical solution for classroom environments where separating students into two sets may not reflect the actual educational setting.

K-Means clustering is computationally efficient and well-suited for large datasets, making it a popular choice in data-driven educational contexts. However, it groups students based on immediate similarities without considering long-term pairing compatibility. As a result, while it can quickly categorize students, it often leads to unstable group formations over time, which may disrupt the learning process [24, 25]. Studies have shown that K-Means lacks the mechanisms needed to maintain the long-term stability of groupings, as it only optimizes for short-term academic or behavioral similarities [25, 26].

Thus, while K-Means is fast and efficient, it is less effective in educational scenarios that require lasting engagement. Gale-Shapley's strength lies in its ability to guarantee stability, although it may require more processing time, particularly with larger groups of students. In contrast, while effective at grouping based on similarity, clustering techniques cannot ensure the long-term stability of pairs, a critical factor in sustaining engagement in educational environments.

The proposed modification to the Gale-Shapley algorithm—incorporating a single list and objective compatibility metrics—offers a robust, efficient, and scalable solution for educational pairing. The modified algorithm ensures stable, compatible pairings that enhance academic and social outcomes by eliminating the need for two distinct lists and integrating metrics such as Jaccard, Cosine, and inverse Euclidean distance. This approach aligns more closely with the natural dynamics of a classroom, offering a practical solution for modern educational environments that balances computational efficiency with the need for stable, meaningful student interactions.

III. Methodology

The proposed pairing model adapts the Gale-Shapley algorithm to operate with a single list of individuals, which is particularly suitable for educational environments with only one list of students. This adaptation responds to the need to develop pairing systems that reflect the specific complexities and dynamics of the academic context. The main innovation lies in modifying the algorithm to deal efficiently with a single list of individuals, finding the pairing based on calculating the compatibility scores between the participants.

The first stage of implementing this system involves calculating compatibility scores. The system determines preferences using these scores, making it suitable for pairing in educational environments with only one list of students. Compatibility scores are calculated based on inverse Euclidean distance, Jaccard similarity, and cosine similarity.

Inverse Euclidean Distance: Measures the inverse distance between two points in multidimensional space, identifying the proximity between students' preferences.

Jaccard Similarity: Compares the similarity and diversity of sample sets, defined as the intersection size divided by the size of the union of the sets.

Cosine Similarity: Measures the cosine of the angle between two vectors, used to calculate similarity regardless of vector magnitude.

These metrics provide different perspectives on similarity or compatibility and are fundamental for calculating compatibility scores in pairing systems, such as the one proposed for educational games.

A. Modification of the Gale-Shapley algorithm

With the compatibility scores calculated, the next step is to run a modified pairing algorithm. The main challenge is adapting the Gale-Shapley algorithm to work with a single list. This modification involves adjusting the algorithm to create stable pairs based on maximizing overall compatibility rather than sticking to bilateral preferences.

Moreover, the adaptation involves running a modified pairing algorithm using the calculated compatibility scores. Traditionally, the Gale-Shapley algorithm deals with two groups and their preference lists [11]. Here, compatibility scores replace bilateral preferences for a single list, creating stable pairs based on overall compatibility. The modified algorithm involves sorting preferences based on compatibility scores, iterating over individuals to form pairs, and ending when no more partnerships exist. This adaptation promotes balanced and efficient pairings, essential for optimizing the learning experience in competitive and collaborative applications.

Our proposal differs from the work presented in Chapter II [16, 17], which use two lists according to the original application of the Gale-Shapley algorithm. In this article, we will modify the algorithm to work with just one list, more accurately reflecting the reality observed in many collaborative and competitive applications in the educational process.

Algorithm 1: Modified Gale-Shapley Matching within a Single Group

Input: A set of individuals $P = \{p_1, p_2, \dots, p_n\}$, Compatibility scores C
Output: A set of pairs S

```
function MODIFIED_GALE_SHAPLEY(P,C)
    Initialize: S ← ∅, FreePeople ← P
    for each p_i ∈ P do
        Preferences[p_i] ← SORT_BY_PREFERENCE(p_i,C)
    end for
    while FreePeople ≠ ∅ do
        p ← CHOOSE(FreePeople)
        for each preferred ∈ Preferences[p] do
            if preferred is free or prefers p over current match then
                if preferred in S then
                    ADD_TO_FREE(S[preferred])
                end if
                S[preferred] ← p
                REMOVE_FROM_FREE(p)
                break
            end if
        end for
    end while
    return S
end function
```

Algorithm 1 has adapted its goal to work with these compatibility scores and a single list.

B. Execution and Assessment

This study uses a rigorous methodology to simulate educational environments with synthetic data, reflecting their diversity. The methodology includes numerical and categorical attributes, facilitating detailed comparative analyses to evaluate the impact of different data types on the effectiveness of pairings. Data normalization ensures comparability and accuracy of compatibility metrics, allowing a fair assessment of similarities and differences between student profiles. The methodology integrates compatibility metrics to calculate participant scores, forming a solid basis for meaningful pairings.

We conduct detailed simulations using synthetic data sets representing student profiles and compatibility metrics to evaluate the algorithm's robustness and adaptability under different conditions. Key metrics such as execution time and success rate measure the algorithm's effectiveness in forming stable pairs and maximizing participant satisfaction. Simulations were conducted for 10, 100, 1,000, and 10,000 individuals to evaluate the algorithm's performance at different scales. This scalable test is crucial for understanding the algorithm's adaptability and efficiency, enabling us to identify performance on varying sizes of datasets.

IV. Results and Discussion

This section analyzes the results of implementing and comparing the modified Gale-Shapley pairing algorithm. We evaluated the algorithm's efficiency and effectiveness using datasets of varying sizes and compositions generated through the Synthetic Data Vault[1] (SDV) package and NumPy[2]. We aimed to explore how different numerical and categorical attributes influence pair formation based on compatibility scores. We used visualizations such as boxplots, bar graphs, and line graphs to examine compatibility distribution, dataset size impact, and interactions among various compatibility metrics.

A. Data Generation

Generating synthetic data is crucial for modeling and analyzing algorithms when accurate data is unavailable, confidential, or incomplete. This section details the methodologies used. The SDV package simulated data with numerical and categorical attributes, while NumPy generated purely numerical data. These approaches allowed us to evaluate the matching algorithms under various conditions, providing insights into their effectiveness and adaptability.

The SDV generates complex synthetic data, simulating multidimensional datasets that reflect real-world diversity. An example of a real dataset can be seen in the work [27]. We used SDV to create datasets with various attributes. This involved normalization and one-hot encoding, converting categorical data into binary matrices for inclusion in compatibility calculations. Min-Max normalization adjusted numerical data values to a standard scale of 0 to 1, ensuring each attribute contributed equally to the compatibility metrics.

In addition to the SDV approach, we used the NumPy library to generate a second dataset focusing exclusively on numerical attributes. This protocol ensured reproducibility and consistency, enabling a direct comparison of the pairing algorithms' performance in a controlled environment.

Table 1 shows a sample of the data before normalization, including numerical and categorical data in text format.

Table 2 displays the normalized data, with numerical data scaled from 0 to 1 and categorical data transformed into Boolean value columns.

Table 1. Sample data generated by SDV before normalization

Interests	Learning Style	Study Preference	Math Proficiency	Reading Proficiency	Science Proficiency	Writing Proficiency	Problem Solving Skills	Social Skills
Arts	Auditory	Group	60.14	29.48	88.61	46.95	17.07	66.94
Literature	Kinesthetic	Group	37.29	87.42	6.38	52.95	8.26	28.33
Arts	Auditory	Individual	36.53	32.82	40.49	40.36	82.33	55.02
Sports	Auditory	Individual	51.51	27.31	85.49	67.81	75.97	85.09
Sciences	Kinesthetic	Individual	38.06	2.52	6.29	58.50	12.88	49.91
Arts	Visual	Group	58.34	16.63	97.60	45.32	24.97	70.70
Sports	Reading / Writing	Group	37.14	66.91	72.74	66.40	7.72	8.13

Table 2. Sample data after normalization and one-hot encoding

Math Proficiency	Reading Proficiency	Science Proficiency	Interests Arts	Interests Literature	Interests Sciences	Interests Sports	Study Preference Group	Study Preference Individual
0.4801	0.2889	0.8911	1	0	0	0	1	0
0.0617	0.8940	0.0539	0	1	0	0	1	0
0.0478	0.3238	0.4013	1	0	0	0	0	1
0.4719	0.3802	0.7282	0	1	0	0	0	1
0.3220	0.2663	0.8594	0	0	0	1	0	1
0.2875	0.6999	0.3815	0	1	0	0	1	0
0.0757	0.0074	0.0530	0	0	1	0	0	1

B. Calculating Compatibility Scores (Compatibility Metric)

We analyzed metric calculation results and the Gale-Shapley algorithm's application of various compatibility metrics, including inverse Euclidean distance, Jaccard similarity, and cosine similarity. These metrics assessed pairing quality and execution time. The choice of compatibility metric also affects the pairing algorithm's runtime, with computational complexity varying for each metric. Therefore, selecting the most appropriate metric based on the dataset and application requirements is crucial for balancing compatibility accuracy and computational efficiency.

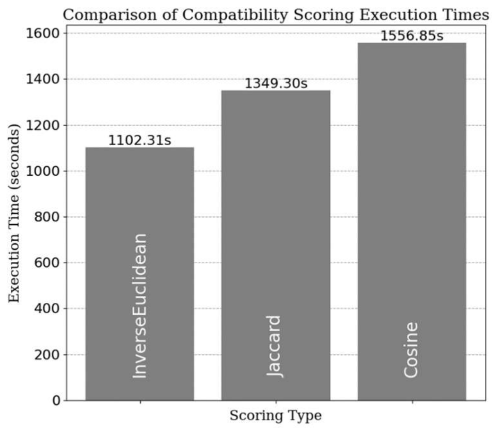

Fig. 1. Graph of execution time for compatibility scoring (numerical data).

[1] https://sdv.dev/
[2] https://numpy.org/

Fig. 1 compares execution times for calculating compatibility metrics with 10,000 individuals using normalized numerical data.

Fig. 2. Graph showing the time it takes to calculate the compatibility score (numerical and categorical data).

Fig. 2 compares execution times for calculating compatibility metrics with 10,000 individuals using numerical and categorical data. Inverse Euclidean compatibility was the fastest in both cases, followed by Jaccard and cosine similarities.

They are implementing different compatibility metrics, which provide insights into how evaluation methods influence matching algorithms. An appropriate metric is critical to achieving high-quality pairings while maintaining computational efficiency.

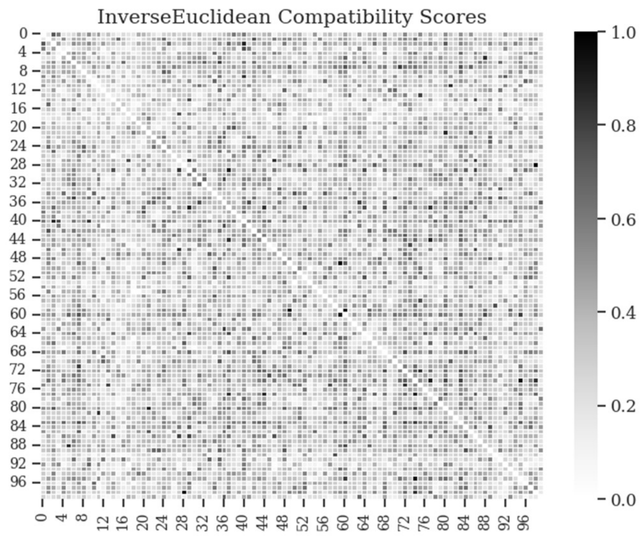

Fig. 3. Heatmap inverse Euclidean compatibility scores.

The heatmap of Fig. 3 displays Inverse Euclidean Normalized Compatibility Scores among students to identify optimal pairings for collaborative learning. High compatibility pairs, indicated by lighter cells, suggest students who are more likely to form stable matches, as they have a higher mutual preference. Conversely, darker cells represent lower compatibility pairs, indicating less optimal matches.

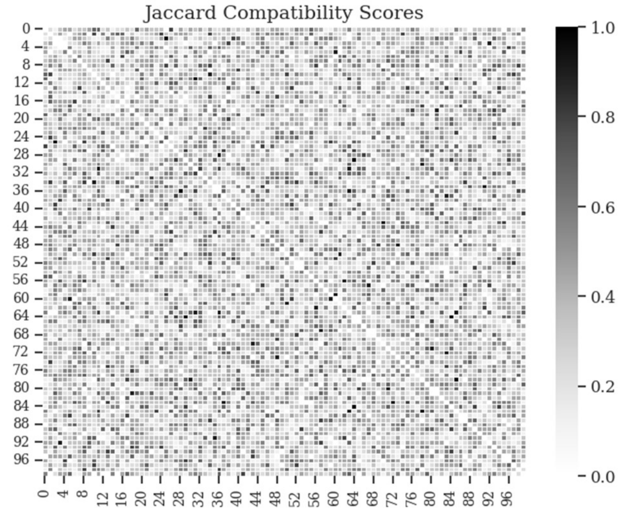

Fig. 4. Heatmap Jaccard compatibility scores.

Similarly, the heatmap of Fig. 4 illustrates Jaccard Normalized Compatibility Scores among students, aimed at identifying optimal pairings for collaborative learning. High compatibility pairs, represented by lighter cells, indicate students with a higher mutual preference, suggesting stable matches. In contrast, darker cells denote lower compatibility, indicating less favorable pairings.

Finally, the heatmap of Fig. 5 shows the pairwise similarity among students based on the cosine similarity metric. The pattern displays distinct horizontal and vertical bands, indicating consistent compatibility clusters where specific groups of students exhibit higher or lower compatibility relative to others. This clustering suggests that certain student groups share similar compatibility profiles, making the cosine metric particularly effective for identifying and forming cohesive groups for collaborative learning. The uniform diagonal line confirms maximal self-similarity, reinforcing the normalization consistency across the dataset.

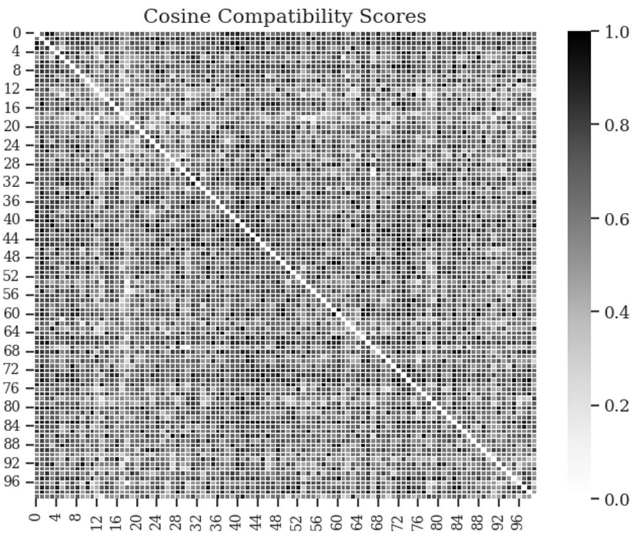

Fig. 5. Heatmap cosine compatibility scores.

The comparison of heatmaps for Inverse Euclidean, Jaccard, and Cosine Normalized Compatibility Scores reveals distinct patterns in student compatibility for collaborative learning. The Inverse Euclidean heatmap shows a diverse compatibility distribution, indicating significant variation in pairwise similarity. The Jaccard heatmap presents a more uniform distribution, suggesting moderate compatibility levels across most pairs, suitable for binary or categorical data. The Cosine heatmap displays clear horizontal and vertical bands, indicating consistent compatibility clusters among specific student groups. These differences highlight the importance of selecting appropriate metrics based on data characteristics and pairing objectives, with each metric providing unique insights into student compatibility.

C. Evaluation of Compatibility Metrics in Matching Algorithms with Numerical Data

We analyzed the results of applying various compatibility metrics—Inverse Euclidean Distance, Jaccard similarity, and Cosine similarity—to the modified Gale-Shapley algorithm using numerical data sets. This analysis provided insights into the nature and interpretation of compatibility scores, especially compared to a random pairing model, from the viewpoint of efficiency (the stability of matched pairs) and efficacy (the computational cost of pairing).

Fig. 6 displays a boxplot comparing pairings generated by the modified Gale-Shapley algorithm using Inverse Euclidean compatibility scores. Groups range from 10 to 10,000 individuals, including a randomly generated group of 10,000. Each box represents the interquartile range (IQR) of compatibility scores, with the median denoted by the line inside the box. Whiskers extend to the most extreme data points not considered outliers.

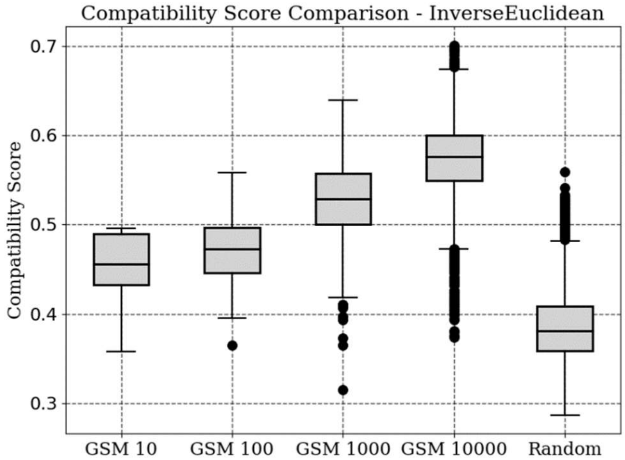

Fig. 6. The boxplot graph illustrates the pairing distributions using inverse Euclidean compatibility scores on the numerical dataset.

Table 3 shows the data used to generate Fig. 6. It indicates that GSM 10 has the lowest average compatibility score (0.4467) with a standard deviation 0.0498. GSM 10000 had the highest average score (0.5724) with a slightly lower standard deviation (0.0418). The randomly generated group had the lowest average score (0.3863) with a similar standard deviation to GSM 10000. Execution time and iterations varied significantly, with GSM 10000 requiring the longest execution time (167.9475 seconds) and iterations (14,579), while GSM 10 had the shortest execution time (0.0002 seconds) and iterations (13).

Table 3. Data of pairing using inverse Euclidean compatibility scores on numerical datasets

	mean	**min**	**25%**	**median**	**75%**	**max**	**execution time**	**iterations**
GSM 10	0.4467	0.3585	0.4329	0.4559	0.4902	0.4961	0.0002	13
GSM 100	0.4734	0.3649	0.4458	0.4733	0.4965	0.5585	0.0141	141
GSM 1000	0.5268	0.3152	0.5007	0.5286	0.5577	0.6394	1.3406	1,448
GSM 10000	0.5724	0.3740	0.5490	0.5762	0.6000	0.7010	167.9475	14,579
Random	0.3863	0.2973	0.3591	0.3831	0.4085	0.6695	0.0037	

Table 4. Data of pairing using Jaccard compatibility scores on numerical datasets

	mean	**min**	**25%**	**median**	**75%**	**max**	**execution time**	**iterations**
GSM 10	0.4444	0.2500	0.2500	0.4375	0.5625	0.6250	0.0002	13
GSM 100	0.4306	0.1250	0.3750	0.4375	0.5000	0.6250	0.0146	122
GSM 1000	0.5257	0.1250	0.5000	0.5000	0.5625	0.6875	1.2894	1,259
GSM 10000	0.6032	0.1875	0.5625	0.6250	0.6250	0.8125	145.2004	12,203
Random	0.1900	0.0000	0.1250	0.1875	0.2500	0.5625	0.0037	

Fig. 7 displays a boxplot of pair distributions using Jaccard index compatibility scores. Like Fig. 6, the graph examines groups from GSM 10 to GSM 10000 and includes a random group of 10,000 individuals.

Table 4 shows that GSM 10 had an average compatibility score of 0.4444 with a high standard deviation of 0.1515. GSM 100 had a slightly lower average of 0.4306 with a standard deviation of 0.0779. GSM 1000 exhibited a higher mean of 0.5257 with a standard deviation of 0.0702. GSM 10000 had the highest average score of 0.6032 with a standard deviation of 0.0600. The random group had a significantly lower average (0.1900) with a broad standard deviation (0.0985). Execution time and iterations varied, with GSM 10000 requiring 145.2004 seconds and 12,203 iterations, while GSM 10 required only 0.0002 seconds and 13 iterations.

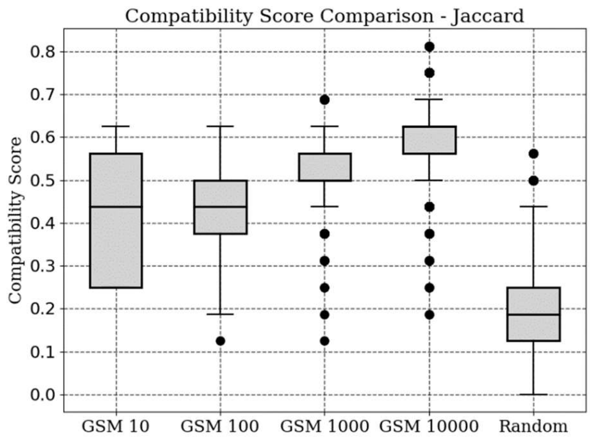

Fig. 7. The boxplot graph displays the pairing distributions using Jaccard compatibility scores on the numerical dataset.

Fig. 8 displays a boxplot of pair distributions using Cosine similarity scores. Groups range from GSM 10 to GSM 10000 and include a random group of 10,000 individuals.

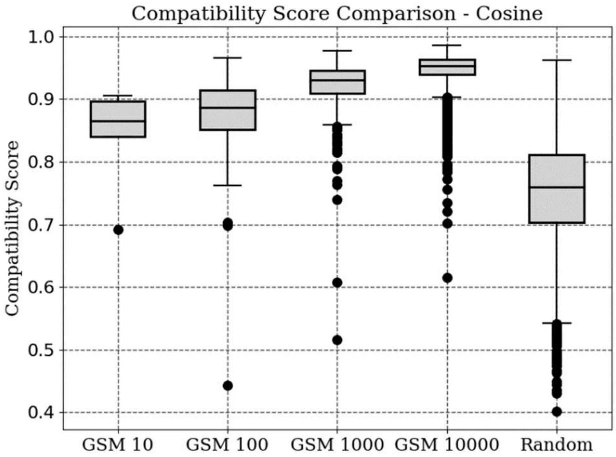

Fig. 8. Boxplot pairing distributions graph using Cosine compatibility scores on the numerical dataset.

Table 5 shows that GSM 10 had an average compatibility score of 0.8398 with a standard deviation of 0.0774. GSM 100 had an average of 0.8682 with a standard deviation of 0.0823. GSM 1000 had an average of 0.9214 with a standard deviation of 0.0408. GSM 10000 had the highest average of 0.9476 with the lowest standard deviation (0.0243). The random group had a lower average (0.7534) with a standard deviation 0.0790. Execution time and iterations increased with group size, with GSM 10000 requiring 152.3835 seconds and 15,004 iterations, while GSM 10 required only 0.0016 seconds and 12 iterations.

Despite numerical differences between metrics, comparing compatibility metrics with random pairing indicates the need to reassess perceived effectiveness. High compatibility scores, mainly with Cosine similarity, do not necessarily imply more efficient pairings. For instance, Cosine similarity generated high scores even for random pairings, raising doubts about the direct correlation between high scores and pairing quality. These results suggest that compatibility values may reflect the characteristics of the metric rather than the actual quality of pairs formed (see Fig. 8).

Table 5. Data of pairing using Cosine compatibility scores on the numerical dataset

	mean	min	25%	median	75%	max	execution time	iterations
GSM 10	0.8398	0.6922	0.8399	0.8648	0.8970	0.9050	0.0016	12
GSM 100	0.8682	0.4430	0.8517	0.8860	0.9143	0.9652	0.1036	158
GSM 1000	0.9214	0.5163	0.9098	0.9302	0.9451	0.9770	3.6140	1,511
GSM 10000	0.9476	0.6156	0.9389	0.9530	0.9629	0.9863	152.3835	15,004
Random	0.7534	0.4018	0.7034	0.7594	0.8109	0.9615	0.0036	

Visual analysis of boxplot graphs (Figs. 6–8) reveals consistent patterns in how metrics respond to dataset size increases. Metrics can differentiate between random pairings and those based on compatibility, emphasizing the importance of evaluating metrics against a random baseline. Qualitative evaluation is crucial alongside quantitative metrics. While metrics provide a comparative basis, researchers should measure practical pairing algorithms by numerical scores and the perceived relevance and satisfaction of the pairs formed. A comparative analysis reveals a trade-off between pairing quality and computational efficiency. The Cosine metric produces the most compatible pairings but requires the most iterations and execution time. The Jaccard metric offers a balanced combination of pairing quality and computational efficiency, making it viable for practical applications requiring high pairing quality without excessive computational time.

Moreover, compared to random pairings, the GSM algorithm improves pairing stability and compatibility across all three metrics (Inverse Euclidean, Jaccard, and Cosine). For Inverse Euclidean and Jaccard, a clear trend of increasing stability and compatibility with higher GSM values suggests that fine-tuning the algorithm parameters can significantly enhance pairing quality. Cosine similarity, on the other hand, exhibits inherently high compatibility scores, indicating its robustness for pairing even at lower GSM values, though improvements are still seen with higher GSM settings.

Also, GSM 10000 consistently produces the most stable and compatible pairs across all metrics, suggesting it is an optimal setting for maximizing pairing stability. The consistently lower and more variable scores for random pairings underscore the importance of using a structured algorithm to achieve stable and compatible student pairings. These insights are crucial for developing and applying algorithms to optimize student pairings for collaborative learning, ensuring that the pairs formed are stable and highly compatible.

These results highlight the importance of carefully selecting compatibility metrics to optimize pairing algorithms. The choice depends on the nature of the data and specific application requirements. The results demonstrate the modified GS algorithm’s adaptability to different metrics and datasets, offering valuable insights for implementing recommender systems and other data-driven pairing applications.

D. Evaluation of Compatibility Metrics: Inclusion of Categorical and Numerical Data

We analyzed the modified Gale-Shapley algorithm using different compatibility metrics—inverse Euclidean distance, Jaccard, and Cosine—on datasets enriched with both categorical and numerical data generated by SDV. Compared to numerical-only data sets, these enriched datasets provide a more complex and representative perspective of human interactions. A detailed analysis, using random pairing performance as a baseline, offers critical insights into the metrics’ effectiveness in capturing compatibility across various attributes.

Fig. 9 shows the boxplot for pair distributions using Inverse Euclidean compatibility scores on datasets containing normalized numerical and categorical attributes. Analysis of

Table 6 and Fig. 9 indicates that GSM 10 has an average score of 0.2894 with a low standard deviation of 0.0172, suggesting homogeneity in compatibility scores. GSM 10000 averages 0.2967 with a low standard deviation of 0.0727, demonstrating consistency even with large data volumes. In contrast, the randomly generated group had a low average of 0.0015 with a standard deviation of 0.0087, indicating almost complete incompatibility. GSM 10000 took 146.5030 seconds for 12,848 iterations.

Comparing Table 6 (mixed data) with Table 3 (numerical-only data), we notice a decrease in compatibility averages when categorical attributes are included. Despite fewer iterations, the execution time for GSM 10000 in the mixed data set is slightly less, suggesting greater processing efficiency despite added complexity. Including categorical attributes significantly impacts compatibility assessment, resulting in lower scores and more significant score variation.

Table 6. Data of pairing using inverse Euclidean compatibility scores on numerical and categorical datasets

	mean	min	25%	median	75%	max	execution time	iterations
GSM 10	0.2894	0.2592	0.2829	0.2953	0.3027	0.3070	0.0002	11
GSM 100	0.2930	0.0708	0.2815	0.2958	0.3318	0.4624	0.0122	125
GSM 1000	0.3003	0.0020	0.2829	0.3027	0.3413	0.4892	1.1398	1,280
GSM 10000	0.2967	0.0003	0.2792	0.3001	0.3381	0.4892	146.5030	12,848
Random	0.0015	0.0001	0.0002	0.0003	0.0007	0.3574	0.0034	

Table 7. Data of pairing using inverse Jaccard compatibility scores on numerical and categorical datasets

	mean	min	25%	median	75%	max	execution time	iterations
GSM 10	0.5765	0.4118	0.5294	0.5882	0.5882	0.7647	0.0003	12
GSM 100	0.6865	0.3529	0.6471	0.7059	0.7059	0.7647	0.0265	122
GSM 1000	0.7799	0.2941	0.7647	0.7647	0.8235	0.9412	1.4173	1,241
GSM 10000	0.8414	0.2941	0.8235	0.8235	0.8824	0.9412	166.7902	12,063
Random	0.4298	0.2353	0.3529	0.4118	0.5294	0.8824	0.0053	

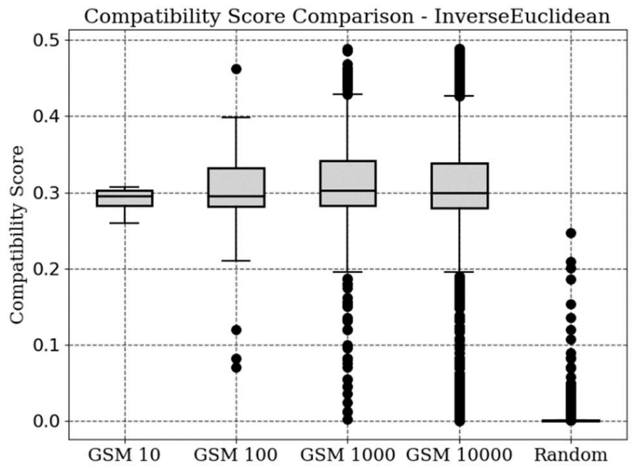

Fig. 9. The boxplot graph shows the pairing distributions using inverse Euclidean compatibility scores on numerical and categorical datasets.

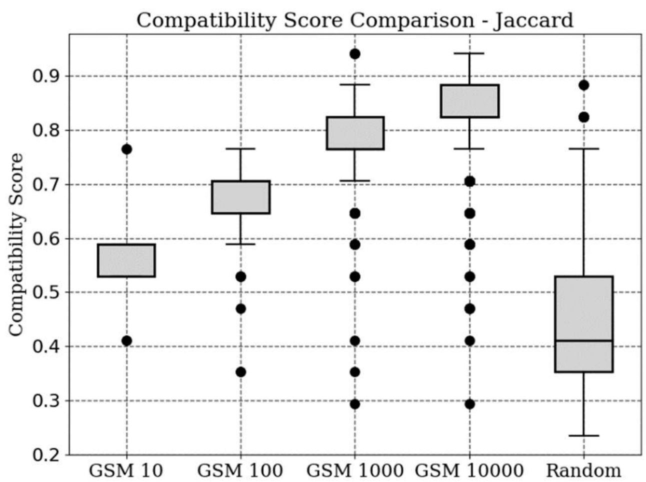

Fig. 10. The boxplot graph illustrates the pairing distributions using Jaccard compatibility scores on numerical and categorical datasets.

Fig. 10 shows the boxplot for pair distributions using Jaccard index compatibility scores. The results indicate significant improvement in pairing quality as the dataset size increases, highlighting the metric's ability to identify true compatibilities in a mixed data environment.

Table 7 shows that GSM 10 has an average compatibility score of 0.5765 with a standard deviation of 0.1141. GSM 100 has an average of 0.6865 with a standard deviation of 0.0802, and GSM 1000 has an average of 0.7799 with a lower standard deviation of 0.0650. GSM 10000 has the highest average compatibility score of 0.8414 with the lowest standard deviation of 0.0496. The random group has a significantly lower average compatibility score of 0.4298 with a standard deviation of 0.1128. Execution times and iterations increase with group size, comparable to Table 6.

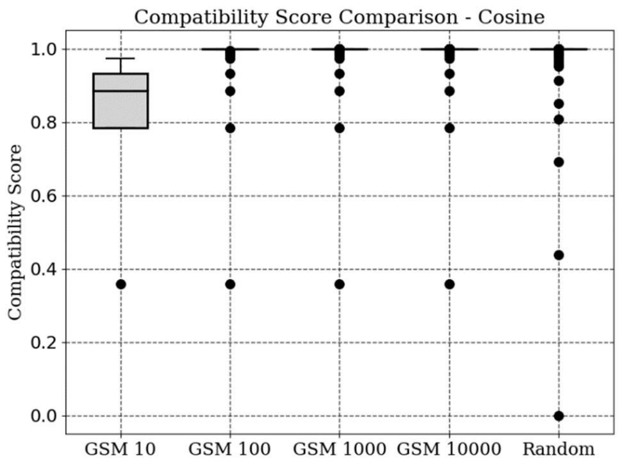

Fig. 11. Boxplot pairing distributions graph using Cosine compatibility scores on numerical and categorical datasets.

Comparing Table 7 (mixed data) with Table 4 (numerical-only data) shows higher averages, indicating that including categorical attributes increases overall compatibility. Both tables show similar results regarding execution time and iterations, with slight increases in the mixed dataset.

Fig. 11 shows the boxplot for pair distributions using Cosine similarity scores. The data in Table 8 indicate the highest average compatibility with the Cosine metric. However, high performance in the random model suggests that the nature of the data might influence high compatibility values. One-hot encoding to normalize categorical data introduces considerable homogeneity and high dimensionality into vectors, potentially inflating compatibility values.

Table 8. Data of pairing using inverse Cosine compatibility scores on numerical and categorical datasets

	mean	min	25%	median	75%	max	execution time	iterations
GSM 10	0.7869	0.3589	0.7850	0.8855	0.9322	0.9728	0.0002	21
GSM 100	0.9774	0.3589	0.9987	0.9997	0.9999	1.0000	0.0152	200
GSM 1000	0.9977	0.3589	1.0000	1.0000	1.0000	1.0000	1.2249	1,607
GSM 10000	0.9998	0.3589	1.0000	1.0000	1.0000	1.0000	163.8606	15,479
Random	0.9995	0.0005	1.0000	1.0000	1.0000	1.0000	0.0036	

Compared to random pairings, the GSM algorithm improves pairing stability and compatibility across all three metrics (Inverse Euclidean, Jaccard, and Cosine). For Inverse Euclidean and Jaccard, a clear trend of increasing stability and compatibility with higher GSM values indicates that fine-tuning the algorithm parameters can significantly enhance pairing quality. The Cosine similarity metric consistently shows high compatibility scores, demonstrating its robustness for pairing even at lower GSM values, though improvements are still seen with higher GSM settings.

GSM 10000 consistently produces the most stable and compatible pairs across all metrics, suggesting it is an optimal setting for maximizing pairing stability. The consistently lower and more variable scores for random pairings underscore the importance of using a structured algorithm to achieve stable and compatible student pairings. These insights are critical for developing and applying algorithms to optimize student pairings for collaborative learning, ensuring that the pairs formed are stable and highly compatible.

Including categorical features generally increases variability and introduces more outliers, but the GSM algorithm still enhances pairing stability and compatibility across all metrics compared to random pairings. Cosine similarity consistently provides the highest compatibility for numerical data alone, while Inverse Euclidean and Jaccard metrics show significant improvements with higher GSM values. When categorical features are added, all metrics show increased variability, but the GSM algorithm's effectiveness in improving stability and compatibility remains evident. GSM 10000 optimizes pairing stability across numerical and combined numerical-categorical datasets. These insights underscore the importance of using structured algorithms for stable and compatible student pairings in collaborative learning contexts.

E. Pairing Analysis: Varying the Number of Individuals.

This section examines how the quality of pairings generated by the modified Gale-Shapley algorithm compares to a random pairing method as the number of individuals varies. We use statistical measures such as mean, standard deviation, minima, maxima, quartiles, and the number of iterations required to achieve a stable pairing to evaluate pairing quality.

The results show a clear improvement in pairing quality as the dataset size increases from 10 to 10,000 individuals, measured by average compatibility. The modified Gale-Shapley algorithm forms higher-quality pairs on average and remains robust and effective across different data scales without compromising pairing integrity.

The standard deviation of compatibility scores increases with dataset size, suggesting greater diversity in pairing quality in larger datasets. Despite this variation, the algorithm still produces high-quality pairings, as evidenced by consistently high maximum compatibility values.

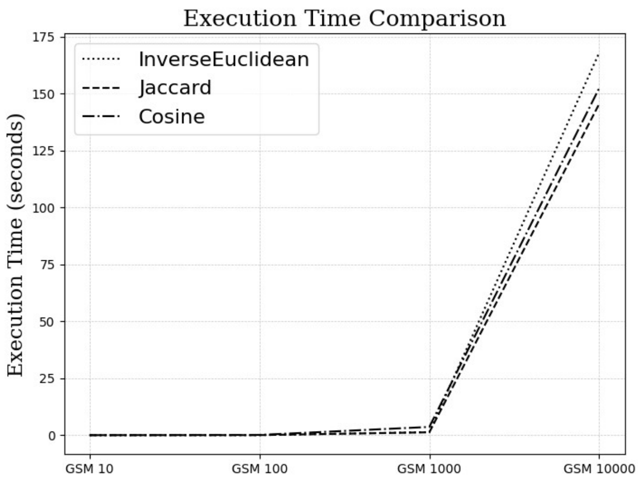

Fig. 12. Comparison of execution time in pair formation for inverse Euclidean, Jaccard and Cosine compatibility metrics.

The execution time comparison chart (Fig. 12) shows the performance of three compatibility metrics, Inverse Euclidean, Jaccard, and Cosine, as the dataset size grows. The execution times for all three metrics increase gradually between GSM 10 and GSM 1000 but rise sharply after GSM 1000, reflecting the increased computational demand at larger scales. While the overall trend is similar across all metrics, Inverse Euclidean shows the steepest increase in execution time, followed closely by Jaccard and Cosine. This indicates that, while all three metrics scale reasonably well with smaller datasets, their performance diverges significantly with larger datasets, highlighting the importance of selecting an appropriate metric based on both dataset size and computational constraints. Despite these increases in execution time, the metrics remain feasible for practical applications involving large datasets, with trade-offs to be considered between precision and efficiency.

The Iterations Comparison chart (Fig. 13) illustrates how the number of iterations required for the modified Gale-Shapley algorithm varies across three compatibility metrics: Inverse Euclidean, Jaccard, and Cosine. Like the execution time analysis, the number of iterations remains relatively stable and low for smaller dataset sizes (GSM 10 to GSM 100), with only a slight increase for GSM 1000. However, after GSM 1000, there is a significant spike in iterations for all three metrics, particularly for the Inverse Euclidean, which requires the most iterations, followed closely by Jaccard and Cosine. This increase in iterations with larger datasets indicates that, while the algorithm is efficient for smaller datasets, its computational cost rises rapidly as the dataset grows, with each metric exhibiting a similar growth pattern. The choice of metric, therefore, becomes crucial in large-scale applications where efficiency in the number of iterations plays a critical role. Despite this increase, the algorithm remains scalable, although the performance differences between metrics become more pronounced with larger datasets.

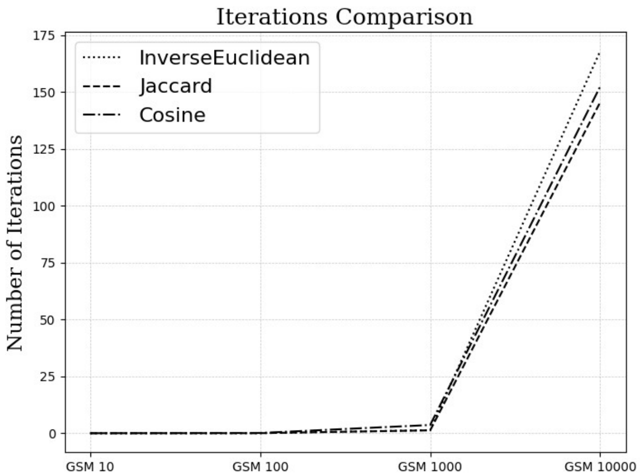

Fig. 13. Comparison of iterations in pair formation for inverse Euclidean, Jaccard and Cosine compatibility metrics.

Analysis of execution time and the number of iterations provides insights into the efficiency and scalability of the modified Gale-Shapley algorithm. While runtime significantly increases with dataset size, the number of iterations increases moderately. These results indicate that the algorithm efficiently finds stable solutions even in large datasets, showing good scalability without excessively compromising efficiency.

The analysis demonstrates that the modified Gale-Shapley algorithm produces high-quality pairings and adapts efficiently to varying dataset sizes. It maintains reasonable efficiency in runtime and iterations, even as data scales increase, proving its applicability in various practical contexts where data-driven pair formation is critical.

V. Conclusions

This study significantly modifies the Gale-Shapley algorithm, adapting it to educational environments requiring a single participant list based on compatibility metrics. The research's main contribution is developing a single-list framework that replaces subjective preferences with objective compatibility scores, improving the stability and quality of pairings in both collaborative and competitive learning settings. Using metrics like inverse Euclidean distance, Jaccard similarity, and cosine similarity, the proposed method ensures stable pairings and optimizes them for educational effectiveness, particularly in promoting SEL.

The critical innovation is adapting the Gale-Shapley algorithm to handle a single list of students, reflecting real-world educational scenarios, such as a classroom. This approach, which relies on objective compatibility scores, provides a practical solution for dynamic classroom environments where predefined preferences are unavailable. This modification enhances pairing stability across different learning contexts, fostering better collaborative and competitive student interactions.

A significant limitation of this study is the reliance on synthetic data to validate the algorithm. While this allows for controlled experimentation and scalability testing, it fails to capture the full complexity of real-world educational environments. Factors such as student emotions, shifting preferences, and classroom dynamics may affect the algorithm's performance in ways synthetic data cannot simulate. Therefore, empirical validation in actual classroom settings is essential to confirm its practical utility and effectiveness under more diverse conditions.

The algorithm scales well in terms of computational complexity, but this scalability comes with a cost. As the number of participants increases—especially with larger datasets of up to 10,000 individuals—execution time and the number of iterations required for stable pairings increase significantly. This is consistent across all compatibility metrics, with inverse Euclidean distance having the highest computational demand.

Future research should focus on several key areas to enhance the application of the modified Gale-Shapley algorithm in real-world educational contexts. First, empirical validation through school field studies, particularly in cooperative and competitive learning scenarios, is crucial. Testing the algorithm in classroom settings or educational game competitions will provide valuable insights into its real-world performance. Further refinement of the compatibility metrics is also necessary, as more sophisticated or hybrid approaches could further improve pairing stability and outcomes. Future studies could explore including behavioral, emotional, or cognitive attributes to enhance pairing success.

Another critical area for future research is the algorithm's adaptability in real-time contexts. Developing educational tools and platforms that integrate this algorithm would be particularly valuable in environments blending cooperative and competitive learning, such as educational games or collaborative problem-solving tasks.

Lastly, testing the algorithm in varied educational settings, including trials in competitive and cooperative educational games, would further demonstrate its applicability. The algorithm could form pairs based on ongoing student performance and interactions, helping bridge the gap between theoretical research and practical applications in diverse educational contexts.

Conflict of Interest

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

LCPJ developed the methodology, implemented the supporting software and algorithms, conducted formal analysis, performed experiments, collected data, and wrote the original draft, also handling data visualization. EG conceptualized the research, conducted experiments, contributed to writing and reviewing the manuscript, and supervised the project while managing resources. LDSM helped formulate the research objectives, supervised the research process, and coordinated project administration. All authors approved the final version of the manuscript.

REFERENCES

[1] C. Cipriano *et al.*, "The state of evidence for social and emotional learning: A contemporary meta-analysis of universal school-based SEL interventions," *Child Development*, vol. 94, no. 5, pp. 1181–1204, Jul. 2023. doi: 10.1111/cdev.13968

[2] J. A. Durlak, R. P. Weissberg, A. B. Dymnicki, R. D. Taylor, and K. B. Schellinger, "The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions," *Child Development*, vol. 82, no. 1, pp. 405–432, Jan. 2011. doi: 10.1111/j.1467-8624.2010.01564.x

[3] J. E. Zins and M. J. Elias, "Social and Emotional Learning: Promoting the development of all students," *Journal of Educational and Psychological Consultation*, vol. 17, no. 2–3, pp. 233–255, Jul. 2007. doi: 10.1080/10474410701413152

[4] Y.-M. Huang, "Exploring students' acceptance of educational computer games from the perspective of learning strategy," *Australasian Journal of Educational Technology*, vol. 35, no. 3, Jun. 2019. doi: 10.14742/ajet.3330

[5] A. F. M. S. Saif, Z. R. Mahayuddin, and A. Shapi'i, "Augmented reality based adaptive and collaborative learning methods for improved primary education towards Fourth Industrial Revolution (IR 4.0)," *International Journal of Advanced Computer Science and Applications/International Journal of Advanced Computer Science & Applications*, vol. 12, no. 6, Jan. 2021. doi: 10.14569/ijacsa.2021.0120672

[6] D. W. Johnson, R. T. Johnson, and K. A. Smith, "Cooperative learning: Improving university instruction by basing practice on validated theory," *Journal on Excellence in College Teaching*, vol. 25, pp. 85–118, Jan. 2014.

[7] A. A. Gokhale, "Collaborative learning enhances critical thinking," *Journal of Technology Education*, vol. 7, no. 1, Sep. 1995. doi: 10.21061/jte.v7i1.a.2

[8] D.-C. Wang and Y.-M. Huang, "Exploring the influence of competition and collaboration on learning performance in digital game-based learning," *Educational Technology Research and Development*, vol. 71, no. 4, pp. 1547–1565, Jun. 2023. doi: 10.1007/s11423-023-10247-8

[9] The Future of Jobs Report 2020, *World Economic Forum*, Oct. 09, 2023.

[10] L. W. Anderson and D. R. Krathwohl, *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*, 2001.

[11] D. Gale and L. S. Shapley, "College admissions and the stability of marriage," *The American Mathematical Monthly*, vol. 69, no. 1, pp. 9–15, Jan. 1962. doi: 10.1080/00029890.1962.11989827

[12] A. E. Roth, T. Sonmez, and M. U. Unver, "Kidney exchange," *The Quarterly Journal of Economics*, vol. 119, no. 2, pp. 457–488, May 2004. doi: 10.1162/0033553041382157

[13] T. Gao, Q. Ye, M. Huang, and Q. Wang, "A systematic model of stable seller–buyer matching with true preference induction in E-commerce platform," *Kybernetes*, Sep. 2022. doi: 10.1108/k-05-2022-0684

[14] A. Abdulkadiroğlu and T. Sönmez, "School Choice: A mechanism Design approach," *The American Economic Review*, vol. 93, no. 3, pp. 729–747, May 2003. doi: 10.1257/000282803322157061

[15] D. Thiruvady, K. Morgan, S. Bedingfield, and A. Nazari, "Allocating students to industry placements using integer programming and ant colony optimisation," *Algorithms*, vol. 14, no. 8, p. 219, Jul. 2021. doi: 10.3390/a14080219

[16] E. Girard, R. Yusri, A. Abusitta, and E. Aïmeur, "An automated stable personalised partner selection for collaborative privacy education," *International Journal of Integrating Technology in Education*, vol. 10, no. 2, pp. 9–22, Jun. 2021. doi: 10.5121/ijite.2021.10202

[17] R. Yusri, A. Abusitta, and E. Aïmeur, "Teens-online: A game theory-based collaborative platform for privacy education," *International Journal of Artificial Intelligence in Education*, vol. 31, no. 4, pp. 726–768, Nov. 2020. doi: 10.1007/s40593-020-00224-0

[18] D. W. Johnson and R. T. Johnson, *Learning Together and Alone: Cooperative, Competitive, and Individualistic Learning*, 1998.

[19] R. E. Slavin, "Research on cooperative learning and achievement: What we know, what we need to know," *Contemporary Educational Psychology*, vol. 21, no. 1, pp. 43–69, Jan. 1996. doi: 10.1006/ceps.1996.0004

[20] S. A. Papert, *Mindstorms: Children, Computers, And Powerful Ideas*, Basic Books, 2020.

[21] A. E. Roth and M. A. O. Sotomayor, "Two-sided matching: A study in game-theoretic modeling and analysis," *The Economic Journal/Economic Journal*, vol. 102, no. 413, p. 975, Jul. 1992. doi: 10.2307/2234601

[22] A. E. Roth, "The evolution of the labor market for medical interns and Residents: a case study in Game Theory," *Journal of Political Economy*, vol. 92, no. 6, pp. 991–1016, Dec. 1984. doi: 10.1086/261272

[23] E. M. Fenoaltea, I. B. Baybusinov, J. Zhao, L. Zhou, and Y.-C. Zhang, "The stable marriage problem: An interdisciplinary review from the physicist's perspective," *Physics Reports*, vol. 917, pp. 1–79, Jun. 2021. doi: 10.1016/j.physrep.2021.03.001

[24] S. Kim *et al.*, "Statistical assessment on student engagement in asynchronous online learning using the k-means clustering algorithm," *Sustainability*, 2023.

[25] A. Dutt *et al.*, "Clustering algorithms applied in educational data mining," *International Journal of Information and Electronics Engineering*, 2015.

[26] D. J. Hall and G. B. Ball, "Clustering educational digital library usage data: A comparison of latent class analysis and k-means algorithms," *Journal of Educational Data Mining*, 2023.

[27] E. Gomede, R. M. Barros, and L. Mendes, "Use of deep multi-target prediction to identify learning styles," *Applied Sciences*, vol. 10, no. 5, p. 1756, Mar. 2020. doi: 10.3390/app10051756