# Collaborative Problem-Solving Process in A Science Serious Game: Exploring Group Action Similarity Trajectory

Jina Kang
Utah State University
Logan, UT 84322
jina.kang@usu.edu

Dongwook An
University of Texas at Austin
Austin, TX 78712
louis.dongwook.an@utexas.edu

Lili Yan
Utah State University
Logan, UT 84322
liliyan@aggiemail.usu.edu

Min Liu
University of Texas at Austin
Austin, TX 78712
mliu@austin.utexas.edu

## ABSTRACT
Collaborative problem-solving (CPS) as a key competency required in the 21st century. There has been an increasing need to understand CPS since it involves not only cognitive but also social processes, and thus its process is difficult to examine. Recent research has highlighted that computer-based learning environments provide an opportunity for students to collaborate with others to solve scientific problems and facilitate their knowledge building process, which can be dynamically tracked within the systems. However, limited research has attempted to identify CPS process captured in the computer-based learning environments designed for supporting CPS. This study therefore aimed to investigate students' CPS process in a serious game, *Alien Rescue*, by analyzing a student's daily tool use action sequence generated in the game. First, we computed a daily gameplay action similarity among students in a group using a similarity coefficient, *Jaccard* (*Jac*). Each group's *Jac* coefficients over the entire gameplay period (i.e. six days over three weeks) were considered as the group action similarity trajectory. The *Jac* coefficient of each day was entered as a single feature (i.e. a total of six features) to conduct a *KmL* cluster analysis that clusters longitudinal data. Three clusters of groups with similar behavior traits (i.e. group action similarity trajectories) were identified. The groups' background information (e.g. solution scores, knowledge gain scores) further provided how the groups' CPS traits can be related to their learning performance.

## Keywords
Collaborative problem-solving, learning process, serious game, *Jaccard* coefficient, *KmL* cluster analysis, science learning.

## 1. INTRODUCTION
Research has highlighted a need for a comprehensive understanding of collaborative problem-solving (CPS), which is regarded as one of the critical competencies of the 21st century skills [10]. OECD [21] recently defined CPS competency as "the capacity of an individual to effectively engage in a process whereby two or more agents attempt to solve a problem by sharing the understanding and effort required to come to a solution and pooling their knowledge, skills and efforts to reach that solution." Computer-based learning environments offer opportunities to monitor collaborative process in order to engage learners in building a shared understanding of a complex problem and support them in knowledge construction process by providing prompts to respond to their learning process or triggering real-time interventions to improve their CPS process [23, 25]. The captured log data including temporal and spatial students' behaviors within the system can reveal emergent patterns that not only reflect individual and group behaviors during CPS activities within the system, but also engage groups with diverse behavior patterns in an effective CPS process accordingly. Despite of these benefits, scanty research has attempted to examine collaboration process captured within a computer-based learning environment designed for supporting CPS. In addition, research on CPS has been mostly conducted by investigating verbal communications (e.g. [13]).

To address this gap, this study aims to investigate groups' CPS process in a serious game, *Alien Rescue*, using an individual student's daily tool use action sequence generated during the entire gameplay period (i.e. a total of six days over three weeks). We used a similarity coefficient, *Jaccard* (*Jac*), to calculate a daily gameplay action similarity among students in a group. *Jac* coefficients of each group over the entire gameplay period, serving as the group's action similarity trajectory, were used to identify patterns of group action similarity trajectory. The findings provide empirical evidences of diverse patterns of CPS process emerged as students engaged in their CPS activities in the serious game. Further, we discuss design considerations of serious games and how our application of the methods can be applied to future studies.

Jina Kang, Dongwook An, Lili Yan and Min Liu "Collaborative problem-solving process in a science serious game: Exploring Group Action Similarity Trajectory" In: *Proceedings of The 12th International Conference on Educational Data Mining (EDM 2019)*, Collin F. Lynch, Agathe Merceron, Michel Desmarais, & Roger Nkambou (eds.) 2019, pp. 336 - 341

## 2. RELEVANT WORK
### 2.1 Collaborative Problem-Solving Process
CPS process has become a field of interest among researchers with the potential to get a better understanding of CPS activities. A review of recent research revealed that CPS phases and synchrony were two topics central to the research on CPS process. Researchers have identified several phases of CPS activities. Informed by

dynamical systems approaches (e.g. [1]), researchers [29] investigated problem-solving phase transition by identifying an entropy peak in transition between the phases including knowledge construction, group problem model, group consensus, and evaluation (see more details in [6]). The entropy peak corresponds to shifts in communication in a problem-solving process within a group. The results showed empirical evidences that groups exhibit phase transitions during their CPS. The concept of initiative has been considered as another way of investigating CPS phases. Howard, Di Eugenio, Jordan, and Katz [11] examined *task initiative* shifts during CPS, which is one type of initiatives and refers to the participation of people in a conversation and their contributions to problem-solving activities during the conversation. They found that group members took *task initiative* when attempting at adding new contents which help the members to advance problem-solving.

Synchrony is another critical factor in understanding CPS process. Synchrony can be developed as group members reach out a shared problem space, which leads to potential transformation or advancement in their problem-solving process. Mercier and Higgins [18] elaborated on the concept of "a joint problem space" [20] and highlighted the importance of a joint understanding of the problem, which can be successfully developed when group members all come to understand the problem that has been worked on. The effectiveness of group work is related to the convergence of the individual members' mental models [4]. Cukurova, Luckin, Millán, and Mavrikis [3] illustrated the significance of synchrony among group members. They provided the evidence of a positive relationship between CPS competence and member synchrony; that is, high competence CPS groups tended to have high levels of member synchrony.

## 2.2 Similarity Coefficients in Serious Games Analytics

A similarity measure is a statistical method to determine how (dis) similar one object is from the other ones by quantifying the similarity or distance between the objects. Typically, the objects being compared can be text strings, audios, images, videos, and navigation sequences, etc. Mathematically, a similarity metric is measured within the range of 0 to 1, indicating two objects are identical (1) and completely different (0). Various applications of similarity measures have been applied in emerging fields of technology, such as audio match and facial recognition. There are five most commonly used similarity measures, namely, Dice, Jaccard (*Jac*), Overlap, Cosine, and the Longest Common Substring coefficients (see details in [15]). In this paper, we used *Jac* coefficient. The use of *n*-gram is an indispensable step to calculate a similarity coefficient, when the directionality or contexts between objects is an important concern [15]. By using *n*-gram, researchers are able to set the sequence of objects before the calculation of similarity coefficients. The *n*-grams are named by the size of the sliding windows used—hence, unigram ($n$ = 1), bigram ($n$ = 2), trigram ($n$ = 3), and fourgram ($n$ = 4) and so on.

A serious game has shown its support to improve learners' CPS performance with the chance to develop problem-solving and collaboration skills and with higher learning motivation (e.g. [25]). As players' actions and behaviors in serious games are considered as the evidence in understanding CPS processes, researchers take serious games as the tool to observe and infer the players' decision-making process (e.g. [15]). Similarity coefficients have been applied to investigate the players' gaming process. Osborn and Mateas [22] defined a (dis) similarity metric for the comparison of players' sequences of actions. They found that the tool *Gamalyzer* (an exploratory visualization of gameplay traces) with the proposed (dis)similarity metric is valid in visualizing the overall strategies of game players. Learning performance in serious games can be quantified with the application of similarity measures to compare the course of action between novice and expert players (e.g. [16]). Loh et al. [15] examined several commonly used similarity measures to determine which measure or combination of measure would be viable in differentiating novice from expert players in serious games. Their findings showed that combining different similarity measures showed stronger predicting abilities than using a single similarity.

# 3. METHODS

## 3.1 Participants

The participants included sixth graders ($n$ = 196) from a middle school in the Southwestern area of the United States. The participants played a serious game, *Alien Rescue*, as a part of science curriculum over three weeks. The teachers encouraged students to group between 2-4 students, but also allowed students to work individually during the gameplay period. There was a total of 70 groups. Each student in a group used their own laptop and solved the problems in collaboration with group members by collecting required information and eliminating planets or moons to find out the most suitable homes for each alien species. In order to investigate students' collaborative problem-solving process, we only included students who worked in a group ($n$ = 156). The students were balanced in terms of gender (77 males and 79 females). At-risk[1] students comprised 51.3% of the sample.

## 3.2 Serious game

*Alien Rescue* (http://alienrescue.edb.utexas.edu) is an open-ended serious game that allows students to discover multiple pathways to solve a problem [9]. In this game, students play in the role of young scientists who are asked to join the United Nations in the effort to rescue six alien species displaced from different places in a distant galaxy by helping them to find new homes in our solar system. Students are engaged in scientific investigations without explicit guidance in their problem-solving process. Students are able to develop a mastery by trying out multiple ways of solving the problem, such in finding evidence, matching information, and formulating rationales. Students develop high-level cognitive skills (i.e. goal setting, hypothesis generation, problem-solving, and self-regulation) while exploring the game environment. The previous studies (e.g. [12, 14]) showed empirical evidences of problem-solving stages within the game; that is, initial exploring and problem identification, background research including gathering and integrating information, hypothesis generation and testing, and solution generation. A set of cognitive tools are provided in the game to support students' problem-solving process (see more details of each tool in [14]). Students are challenged to identify relevant information of the solar system by using in-game cognitive tools and match the information with each alien's needs and characteristics. To solve the complex and ill-structured problem, students need to use the tools strategically. Students get access to the cognitive tools through a two-layer interface. Tools in first layer

[1] The school identified students as being at-risk of dropping out of a school by the state-defined criteria including low-performance on an assessment instrument and limited English proficiency [27].

can be accessed one at a time while the six tools in the second layer can be used anytime overlaid with other tools (see Figure 1).

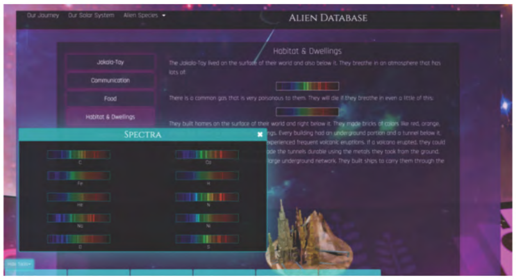


**Figure 1. Alien Database overlaid with Spectra**

## 3.3 Data sources

### *3.3.1 Performance scores*

Before and after gameplay, an individual student's comprehension of factual and applied scientific knowledge introduced in the game was measured using a Space Science Knowledge Test (SSKT). SSKT consists of twenty-four multiple choice items (Cronbach's alpha = 0.77), which score ranges between 0 and 24 (1 point for each item) logged in the system. An individual student's SSKT gain score was calculated by subtracting the pretest score from the posttest score. Then, each group's average gain score was calculated (i.e. total gain scores / a number of students in a group), which is considered as each group's after-game performance.

In addition, the game logs a student's written recommendation(s) for each alien, in which they must indicate an appropriate home for each species and provide a rationale. Students can submit multiple recommendations for each alien species, which reveals the results of students' problem-solving processes—that is, justifications of their solutions using the gathered data during the gameplay. The solutions were evaluated using an 8-point rubric used in previous studies (see more details in [2]) in terms of the correctness of the solution and the number of reasons to the selected home. Each group's average solution score is considered as the group's in-game performance.

### *3.3.2 Gameplay data*

The gameplay data—that is, the user-generated data derived directly from students' actions within the game—were used to identify students' navigation patterns as they engaged in *Alien Rescue*. The game logs every action as each student interacts with the environment. The gameplay data contains a student identifier, a cognitive tool that the student accessed, a type of action (e.g., open, close, click), an additional note on student's interactions, and a timestamp for each action (see an example of data in Table 1). "Open" indicates a student opens a tool, while "Click" indicates a student clicks a submenu of the tool.

**Table 1. Example of A Student's Navigation Data**

| Tool | Action | Notes | Timestamp |
|---|---|---|---|
| Probe Design | Open | | 5/17 10:33:19 |
| Solar System | Click system | Mercury | 5/17 10:36:48 |
| Concepts | Open | | 5/17 10:48:00 |
| Concepts | Close | | 5/17 10:49:06 |

## 3.4 Analysis

### *3.4.1 Group action similarity using a Jaccard coefficient*

We computed a gameplay action similarity between students in a group with a *Jac* coefficient (see 2.2). In order to calculate the similarity of students' navigation traces for each day, we cleaned and transformed each student's navigation data into a 'bag of words.' For example, assuming that on Day1, one student's navigation is represented by *string* A = "*Probe Design* Open, *Solar System* Click *Mercury*, *Concepts* Open, …", and the other student's data is expressed by *string* B = "*Solar System* Open, *Solar System* Click *Mercury*, *Solar System* Click *Venus*, …", we can obtain the intersection set of A and B ($A \cap B$) and union set of $A$ and $B$ ($A \cup B$). The *Jac* coefficient is therefore the length of intersection set over the length of the union set. In order to check if there is a directionality between students' actions, we also applied a bigram setting to the navigation sequence and calculated a *Jac* coefficient. A bigram sequence was obtained using a 'sliding window' of size 2. We conducted a descriptive analysis to compare the distributions of the unigram with bigram *Jac* coefficients of all groups. As shown in Figure 2, the groups' unigram *Jac* coefficients were overall normally distributed, while the bigram *Jac* coefficients showed highly skewed to zero (i.e. small variance). Therefore, we decided to use a unigram *Jac* coefficient for this study. As Loh et al. [15] suggested using a larger *n*-gram, when any contextual relationship between actions are critical, we included a type of action (see 3.3.2) and split the data into each day to further consider the context and directionality.

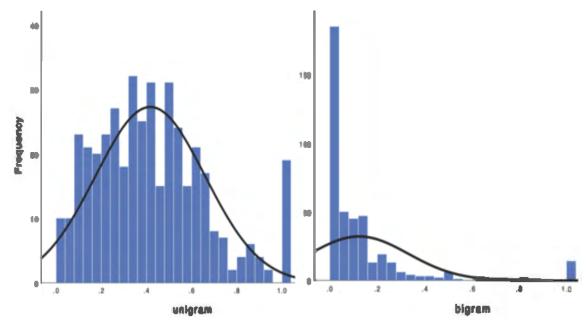


**Figure 2. Histograms of unigram and bigram *Jac* coefficients**

*Note.* Unigram ($M$ = .42, $SD$ = .24), Bigram ($M$ = .12, $SD$ = .21)

### *3.4.2 Clustering analysis for longitudinal data*

A *Jac* coefficient of each day was entered as a single feature (i.e. a total of six features) to conduct a cluster analysis to identify the potential clusters of collaborative groups with similar behavior traits (i.e. group action similarity trajectory). Six *Jac* coefficients of one group can be seen as the action similarity trajectory of the group over the gameplay. To handle such trajectory data, we used a *KmL* package in R, which is a new implementation of *k*-means designed to analyze longitudinal data [8]. One common problem in longitudinal studies is missing data (e.g. [17]). While *k*-means is unable to handle missing values and normally excludes missing data, *KmL* provides diverse imputation methods (e.g. linear interpolation, copyMean; see more details [7]) to deal with different types of missing values including intermittent missing data (data missing in the middle of a trajectory) and monotone missing data (data missing either at the beginning or end) [19]. We were unable to calculate a coefficient when there was only one student in a group logged in the game on a certain day. There were such missing values found randomly over six gameplay days. Therefore, we used linear interpolation (Bisector) to handle missing values of the *Jac* dataset, since the method considers not only a local intermediate line (not just sensitive to first or last values), but also a bisector between a global and local lines.

*KmL* provides methods to define starting conditions and an optimal number of clusters and an easy way to run several *k*-means. *KmL* transforms longitudinal data into an object called 'ClusterLongData.' Once the object has been created, *KmL* runs *k*-means several times and stores all the clusters that the algorithm finds over each iteration of finding an optimal partition in the object. *KmL* also offers a tool that can visualize partitions, in which researchers can make a decision on the best partition by comparing different criteria including Calinski & Harabatz, Ray & Turi, and Davies & Bouldin (e.g. [7]). In this study, 3 clusters were suggested as the optimal number of clusters.

The data failed the major assumption of the one-way ANOVA (i.e. the non-normally distribution assumption). We thus conducted a non-parametric test, Kruskal-Wallis H test, to confirm a statistical significance of the group action similarity trajectories between clusters since a cluster analysis can only reveal the latent cluster patterns. In addition, the cluster patterns were visualized for deeper understanding of group action similarity trajectories in each cluster.

## 4. RESULTS

As shown in Table 2, the average values of daily *Jac* coefficients as exhibited by the three clusters of groups achieved the level of significance ($\chi^2$), indicating the groups were well-partitioned into each cluster. Kruskal-Wallis H tests overall showed that there were statistically significant differences between the mean ranks of daily *Jac* coefficients at least in one pair of clusters. Dunn's pairwise tests for each *Jac* coefficient were carried out for the three pairs of clusters (i.e. Cluster 1 & Cluster 2, Cluster 2 & Cluster 3, and Cluster 1 & Cluster 3). We further examined the background of the groups in each cluster including the average SSKT gain score and the number of groups who submitted at least one solution. To further investigate potential patterns between the clusters, line charts of action similarity trajectories grouped by each cluster including a similarity trajectory trend were derived (see Figure 3). As shown in Figure 3, the line charts indicated that the similarity trajectory trends of the groups in each cluster were distinctively different.

Approximately 40% of the groups are centered in Cluster 1, and the mean ranks of *Jac* coefficients were overall lower than those of the other two clusters. As shown in Figure 3, this cluster's overall similarity trajectory decreased slightly. This cluster showed the lowest solution submission rate. These groups achieved the lower average solution scores and SSKT gain scores than the other clusters.

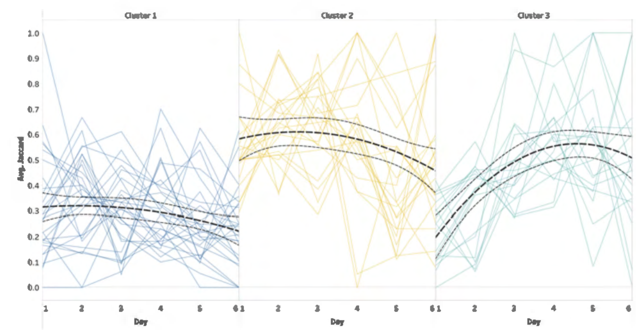


**Figure 3. Similarity Trajectories of Groups in Each Cluster**

*Note.* The dashed line shows a polynomial trend model of degree 2 computed for 'Average of *Jac*' given 'Day' (dotted lines for a confidence band).

About 30% of the groups are in Cluster 2, and the groups exhibited overall the highest mean ranks of *Jac* coefficients except for Day 5. Similar to Cluster 1, the groups in Cluster 2 exhibited a decreasing similarity trajectory trend toward the end of gameplay, but maintained high similarity in their group actions during the first few days. The groups in this cluster showed the highest solution submission rate. Both their average solution and SSKT gain scores showed that their average performance was close to the mean performance of all clusters.

Lastly, the rest of the groups are centered in Cluster 3. The groups in Cluster 3 exhibited an increasing similarity trajectory trend with the peak on Day 5. It is worth noting that their *Jac* coefficients during the first two days were recorded as the lowest mean ranks among all clusters. The solution submission rate of Cluster 3 is close to the average solution submission rate across all clusters (31.58%). However, the groups who submitted at least one solution performed better at their solution submissions than other clusters. Additionally, the groups in this cluster performed the best at their SSKT gain score.

**Table 2. Cluster Membership Description**

| **Cluster** | **Group SSKT Gain score (Mean)** | **[a]No. of groups with solution(s)** | ***Jac* Day1** | ***Jac* Day2** | ***Jac* Day3** | ***Jac* Day4** | ***Jac* Day5** | ***Jac* Day6** | ***Jac* (Mean)** |
|---|---|---|---|---|---|---|---|---|---|
| C1 ($n$ = 29, 41.43%) | 1.727 | 6 (20.69%, [b]$M_{solution}$ = 2.00) | 0.3057 [d](30.12) | 0.3398 (27.10) | 0.2901 (20.26) | 0.3079 (20.55) | 0.2503 (23.98) | 0.2256 (22.21) | 0.2866 |
| C2 ($n$ = 22, 31.43%) | 2.682 | 9 (40.91%, $M_{solution}$ = 3.67) | 0.5473 (53.84) | 0.6427 (55.41) | 0.6420 (52.55) | 0.5982 (46.64) | 0.4149 (38.02) | 0.5170 (45.18) | 0.5603 |
| C3 ($n$ = 19, 27.14%) | 3.412 | 6 (31.58%, $M_{solution}$ = 4.00) | 0.2218 (22.47) | 0.3213 (25.26) | 0.4961 (39.03) | 0.5785 (45.42) | 0.5609 (50.16) | 0.4963 (44.58) | 0.4458 |
| [c]$\chi^2$ | | | $27.73^{***}$ | $30.80^{***}$ | $32.29^{***}$ | $26.77^{***}$ | $19.49^{***}$ | $21.18^{***}$ | |

*Note.* [a]The number of groups submitted at least one solution (Percentage of the groups in each cluster); [b]Average of the total solution scores of groups in each cluster; [c]Kruskal-Wallis H test results, $^{***}p$ <0.001; [d]Mean ranks

## 5. DISCUSSION

We applied a *Jac* coefficient to investigate CPS process within a serious game, *Alien Rescue*: that is, (1) to compute a daily gameplay action similarity among students in a group and (2) to identify action similarity trajectory patterns across groups using each group's similarity trajectory.

Although the groups in Clusters 1 and 2 exhibited overall a decreasing similarity trend, Cluster 1 showed relatively lower action similarities over the entire gameplay, indicating the group members did not use the same tools. The group members seemed to maintain the individual tool use tendency over time. Both Clusters 2 and 3 showed relatively higher *Jac* coefficients and higher in-game (i.e. solution score) and after-game (i.e. knowledge gain score) learning performances, which, together with the chi-square test, can be most likely seen there might be a potential positive relationship between a *Jac* coefficient and learning performance. However, the two clusters' *Jac* coefficients followed a different pattern; that is, the *Jac* coefficients of Cluster 3 have risen considerably over the gameplay period, and the group in Cluster 3 started off with the lowest action similarity among all clusters. Additionally, the results of Dunn's pairwise tests showed the *Jac* coefficients were significantly different ($p < .001$) between Cluster 2 and Cluster 3 only during the first two days. Research has shown the importance of the convergence of individual's mental models in CPS [4] and the positive association between CPS competence and a level of member synchrony [3]. The findings therefore suggest that, during the early gameplay days, the group members in Cluster 3 came to successfully understand the problem and engaged in their collective cognitive process. This further supports the fact that a group action similarity trajectory can be an indicator of the process of developing shared problem space between group members [18].

We applied *n*-gram to compute a *Jac* similarity coefficient: unigram and bigram. Compared with the groups' unigram *Jac* coefficients, the bigram *Jac* coefficients showed a small variance (i.e. highly skewed to zero). In this study, a unigram *Jac* coefficient is therefore a viable way in understanding different levels of group collaboration in this serious game. Research on serious games analytics highlighted the use of *n*-gram would be critical to understand directionality and contexts between events or actions [15, 28]. Since a unigram can possibly ignore the context and directionality of actions, we included a type of action (i.e. click a sub-menu in each tool) to further consider the context and split the data into each day to preserve the directionality when computing a *Jac* coefficient. Such modification is needed when applying *n*-grams to different purposes of study. In addition, we applied a bigram to further examine the frequent sequences of groups in each cluster.

*KmL*, a cluster analysis for longitudinal data, has been often used in scientific disciplines such in medical research [7, 8]. The *KmL* clustering results in this current study showed remarkable differences of the groups' action similarity trajectories among three clusters, which indicate different patterns of CPS process in the serious game. In particular, the positive action similarity growth of Cluster 3 demonstrated that they developed a shared understanding of the problem during the early gameplay days, which has been considered as a critical process of successful collaboration in CPS activities [5, 23]. It is confirmed by the fact that the learning performance of group members of Cluster 3 was higher than that of the groups in other clusters, indicating their experience throughout the CPS process was successfully transformed to their knowledge gain. The findings highlight the importance of providing guidance for students who tend to work independently (i.e. Cluster 1) or who may simply replicate actions of other students in the group (i.e. Cluster 2) to engage in the process of developing a shared problem space. The results further inform design considerations of serious games that support CPS: for example, providing prompts with explicit inquiries, in which a group can be engaged in the successful CPS process grounded on the group's achievement of a shared understanding of a given problem. Taken together, this study confirms *KmL* as a promising method to examine features at different time points generated from gameplay data, which can be seen as an action trajectory that provides insights into CPS process in serious games.

Our work has limitations that should be addressed in future studies. First, the dataset is small and was collected at one middle school with little diversity; for example, 51.3% of the sample was labeled as at-risk, which may not be applicable in other schools with different settings. Second, understanding CPS process is critical, but challenging particularity in an open-ended learning environment like *Alien Rescue*. This work therefore should be expanded to include additional data such as video or audio recordings to capture group conversations and actions to provide robust evidence for the findings of this study. Third, our method of clustering group action trajectory patterns using the *KmL* clustering together with a *Jac* coefficient showed promising evidences to understand students' CPS process. However, due to the small sample size, this may need to be further explored at a larger scale. We are currently employing integrated analytical methods to better understand CPS process using such as mixture latent growth curve model to compare the cluster memberships with the results from *KmL*, and multilevel modeling to examine the relative influence of teacher (i.e. two teachers in the middle school) on the action similarity trajectories of the groups.

## 6. CONCLUSION

This study used a student's daily tool use action sequence generated in a serious game, *Alien Rescue*, to investigate the students' CPS process. We applied a similarity coefficient, *Jac*, to identify a group action similarity trajectory. The *KmL* clustering analysis discovered unique clusters of groups with similar group action trajectories, the membership of which further provided how CPS traits can be related to their learning performance. Each cluster's characteristics shed light on deriving design considerations to promote students' positive collaboration experience during CPS activities within serious games, and to engage teachers in facilitating students' effective CPS process. Lastly, the advantages and limitations of the methods employed in this study point toward the need for continued research on exploring potential analytical methods and scaling up the sample size to include more diverse population.

## 7. REFERENCES


[1] Angus, D., Smith, A. E., and Wiles, J. 2012. Human communication as coupled time series: Quantifying multi-participant recurrence. *IEEE Transactions on Audio, Speech, and Language Processing*, *20*, 6 (Aug. 2012), 1795-1807. DOI= https://doi.org/10.1109/TASL.2012.2189566.

[2] Bogard, T., Liu, M., and Chiang, Y. H. 2013. Thresholds of knowledge development in complex problem solving: A multiple-case study of advanced learners' cognitive processes. *Etr. & Educ. Tech. Res. 61*,3 (Jun. 2013), 465-503. DOI= https://doi.org/10.1007/s11423-013-9295-4.

[3] Cukurova, M., Luckin, R., Millán, E., and Mavrikis, M. 2018. The NISPI framework: Analysing collaborative problem-solving from students' physical interactions. *Computers & Education*, 116 (Jan 2018), 93-109. DOI= https://doi.org/10.1016/j.compedu.2017.08.007.

[4] DeFranco, J. F., Neill, C. J., and Clariana, R. B. 2011. A cognitive collaborative model to improve performance in engineering teams—A study of team outcomes and mental model sharing. *Syst. Eng.* 14, 3 (Oct. 2011), 267-278. DOI= https://doi.org/10.1002/sys.20178.

[5] Dillenbourg, P. 1999. What do you mean by collaborative learning?. P. Dillenbourg. *Collaborative- learning: Cognitive and Computational Approaches*., Oxford: Elsevier, pp.1-19.

[6] Fiore, S. M., Rosen, M. A., Smith-Jentsch, K. A., Salas, E., Letsky, M., and Warner, N. 2010. Toward an understanding of macrocognition in teams: Predicting processes in complex collaborative contexts. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 52, 2 (Jul. 2010), 203-224. DOI= https://doi.org/10.1177/0018720810369807.

[7] Genolini, C., Alacoque, X., Sentenac, M., and Arnaud, C. 2015. kml and kml3d: R packages to cluster longitudinal data. *J. Stat. Softw. 65*(4), 1-34. DOI= https://doi.org/10.18637/jss.v065.i04.

[8] Genolini, C., and Falissard, B. 2011. KmL: A package to cluster longitudinal data. *Comput. Meth. Prog. Bio*. 104, 3 (Dec. 2011), 112-121. DOI= https://doi.org/10.1016/j.cmpb.2011.05.008.

[9] Glaser, R. 1991. The maturing of the relationship between the science of learning and cognition and educational practice. *Learning and Instruction,* 1, 2 (1991), 129–144. DOI= https://doi.org/10.1016/0959-4752(91)90023-2.

[10] Griffin, P., McGaw, B., and Care, E. 2012. *Assessment and teaching of 21st century skills: Methods and Approach.* New York, NY: Springer. DOI= https://doi.org/10.1007/978-94-007-2324-5.

[11] Howard, C., Di Eugenio, B., Jordan, P., and Katz, S. 2017. Exploring initiative as a signal of knowledge co-construction during collaborative problem solving. *Cognitive sci.* 41, 6 (Nov. 2017), 1422-1449. DOI= https://doi.org/10.1111/cogs.12415.

[12] Kang, J., Liu, M., and Qu, W. 2017. Using gameplay data to examine learning behavior patterns in a serious game. *Comput. Hum. Behav*, 72 (Jul. 2017), 757-770. DOI= https://doi.org/10.1016/j.chb.2016.09.062.

[13] Keyton, J., Beck, S. J., and Asbury, M. B. 2010. Macrocognition: a communication perspective. *Theoretical Issues in Ergonomics Science*, 11, 4 (Jun 2010), 272-286. DOI= https://doi.org/10.1080/14639221003729136.

[14] Liu, M., and Bera, S. 2005. An analysis of cognitive tool use patterns in a hypermedia learning environment. *Educational Technology Research and Development, 53*, 1 (Mar. 2005), 5–21.DOI= https://doi.org/10.1007/BF02504854.

[15] Loh, C. S., Li, I. H., and Sheng, Y. 2016. Comparison of similarity measures to differentiate players' actions and decision-making profiles in serious games analytics. *Comput. Hum. Behav. 64*, 562-574. DOI= https://doi.org/10.1016/j.chb.2016.07.024.

[16] Loh, C. S., and Sheng, Y. 2015. Measuring the (dis-)similarity between expert and novice behaviors as serious games analytics. *Education and Information Technologies*, 20, 1 (Nov. 2015), 5–19. DOI= http://doi.org/10.1007/s10639-013-9263-y.

[17] Mallinckrod, C. H., Lane, P. W., Schnell, D., Peng, Y., and Mancuso, J. P. 2008. Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials. *Drug Inf. J*. 42, 4 (Jul.2008), 303-319. DOI= https://doi.org/10.1177/009286150804200402.

[18] Mercier, E., and Higgins, S. 2014. Creating joint representations of collaborative problem solving with multi-touch technology. *J. Comput. Assist. Lear*. *30*, 6 (Feb. 2014), 497-510. DOI= https://doi.org/10.1111/jcal.12052.

[19] Molenberghs, G., Thijs, H., Jansen, I., Beunckens, C., Kenward, M. G., Mallinckrodt, C., and Carroll, R. J. 2004. Analyzing incomplete longitudinal clinical trial data. *Biostatistics*, *5*, 3 (Jul. 2004), 445-464. DOI= https://doi.org/10.1093/biostatistics/kxh001.

[20] Newell, A., and Simon, H. A. 1972. *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

[21] Organisation for Economic Co-operation and Development. (2017). PISA 2015: Draft collaborative problem solving framework. Retrieved from https://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf.

[22] Osborn, J. C., and Mateas, M. 2014. A game-independent play trace dissimilarity metric. In FDG (April).

[23] Roschelle, J., and Teasley, S. D. 1995. The construction of shared knowledge in collaborative problem solving. In *Computer supported collaborative learning* (pp. 69-97). Springer, Berlin, Heidelberg.

[24] Saab, N., van Joolingen, W., and van Hout-Wolters, B. 2012. Support of the collaborative inquiry learning process: Influence of support on task and team regulation. *Metacog. Learn*, 7, 1 (Mar. 2012), 7-23. DOI= https://doi.org/10.1007/s11409-011-9068-6.

[25] Sánchez, J., and Olivares, R. 2011. Problem solving and collaboration using mobile serious games. *Computers & Education*, 57, 3 (Nov. 2011), 1943-1952. DOI= https://doi.org/10.1016/j.compedu.2011.04.012.

[26] Stewart, A. E., Keirn, Z. A., and D'Mello, S. K. 2018 October. Multimodal Modeling of Coordination and Coregulation Patterns in Speech Rate during Triadic Collaborative Problem Solving. In *Proceedings of the 2018 on International Conference on Multimodal Interaction* (pp. 21-30). ACM. DOI= https://doi.org/10.1145/3242969.3242989.

[27] Texas Education Agency. (2017).

[28] van der Loo, M. P. 2014. The stringdist package for approximate string matching. *The R Journal*, *6*, 1 (2014), 111-122. https://doi.org/10.32614/RJ-2014-011.

[29] Wiltshire, T. J., Butner, J. E., and Fiore, S. M. 2018. Problem-Solving Phase Transitions During Team Collaboration. *Cognitive Sci.*, 42, 1(Jan. 2018), 129-167. DOI= https://doi.org/10.1111/cogs.12482.