A Clustering-based Grouping Model for Enhancing Collaborative Learning

Yulei Pang[a], Feiya Xiao[b], Huaying Wang[c], Xiaozhen Xue[d]
[a] Department of Mathematics, Southern Connecticut State University, USA, pangy1@southernct.edu
[b] Department of Educational Psychology and Leadership, Texas Tech University, USA, feiya.xiao@ttu.edu
[c] LMIB-School of Mathematics and Systems Science, Beihang University, China, matheson@smss.buaa.edu.cn
[d] Department of Computer Science, Texas Tech Univeristy, USA, xiaozhen.xue@ttu.edu

***Abstract*—Groupwork is widely used in tertiary institutions due to the considerable advantages of collaborative learning. Previous studies indicated that the group diversity had positive influence on the groupwork achievement. Therefore, how to achieve diversity within a group effectively and automatically is an interesting question. In this paper we propose a novel clustering-based grouping model. The proposed technique first employs balanced *K*-means algorithm to divide the students into several size-balanced clusters, such that the students within the same cluster are more similar (in some sense) to each other than to those in other clusters; then adopts one-sample-each-cluster strategy to construct the groups.[1] We evaluated the proposed technique based on two small-scale case studies. The result observed may indicate that the clustering-based grouping model is feasible and effective.**

***Keywords*—balanced *K*-means, group diversity, collaborative learning**

I. Introduction

The importance and benefit of collaborative learning have been recognized and widely accepted [1]. Collaborative learning in the form of learn/work group encourages students to reach out to each other to solve problems, thus building cooperation and communication skills. Also, individuals in the same group share resources including task-related knowledge, potentially useful information, possible solutions or ideas, evaluation of each other's work and so on, so that collaborative learning leads to deeper understanding of the materials [2]. In order to improve the effectiveness of collaborative learning and better apply it into teaching practice, a number of researchers focus on the theory of workgroup organizations, and recognize group diversity as a major factor.

Group diversity is known as the difference amongst the group members with respect to their backgrounds, abilities, and characteristics [3]. Although it is still an under-explored research question in education and psychology literature how group diversity affects groupwork achievement, a considerable amount of studies provide evidences that diversity positively impacts learning outcomes in work groups. An immediate question that arises is how to implement the diversity effectively and automatically in work groups. The question appears easy to solve when only single or a couple of diverse features are taken into consideration. However, recent studies explored a lot more features which have significant influence on the groupwork achievement, thus making the manual implementation of group diversity complicated. Therefore, it is needed to propose techniques that organize groups automatically and effectively.

We may confront challenges when implementing diversity (dissimilarity) within groups by using existing techniques. However, converting the problem into its opposite proposition (i.e., accomplishing similarity in each cluster) is feasible. First, we apply the balanced *K*-means algorithm[4] to cluster all the observations, which are students in this scenario, such that all elements in the same cluster share the strong likeness. Next, we construct a group by randomly choosing one student from each cluster, thus allowing each group approach less similarity. This paper propose a novel cluster-based technique to group students so as to implement the group diversity automatically and effectively. The contributions of this paper are as follows:

- Introducing a grouping model for the purpose of collaborative learning;
- Adopting the balanced *K*-means algorithm to cluster students according to different diversity criteria;
- Evaluating the performance of the proposed technique based on two small-scale case studies.

The rest of this paper is organized as follows. Section II describes the background knowledge of collaborative learning and group diversity, and reviews the balanced *K*-means algorithm that are referred to in this paper. Section III provides a detailed explanation of the proposed technique. In Section IV, we evaluate our technique based on a number of experiments. Section V concludes our study and discusses possible future research directions.

II. Background

This section first presents the background knowledge including advantages of groupwork and the influence of diversity. It then reviews the balanced *K*-means algorithms that are adopted for the grouping model.

A. Groupwork and Diversity

Collaborative learning refers to a situation where two or more people learn or attempt to learn something together [1], and it is a common held belief that groupwork is the most general form of collaborative learning. The importance

[1]It is important to note that cluster and group have different meaning in this paper. Cluster is a notation in machine learning where objects in the same cluster are more similar, whereas, group is defined as the work/study team we are trying to construct.

978-1-4799-7415-3/14 $31.00 © 2014 IEEE
DOI 10.1109/ICMLA.2014.94

of groupwork has been widely emphasized [5], and a bunch of researchers have discussed and presented the principal ground in detail[6][7][8]. Davis [9] conducted a survey on the advantages of groupwork and he summarized them into 5 categories: 1) groupwork promotes deep as opposed to surface learning; 2) groupwork promotes active as opposed to passive learning; 3) groupwork promotes experiential learning; 4) groupwork can be justified on the grounds of promoting the construction of knowledge; 5) groupwork is also claimed to be an authentic form of assessment in terms of a student's later employability.

Due to the advantages mentioned above, groupwork is widely used in education institutions. Correspondingly, the group organization theory is becoming more and more important. In group organization theory, groupwork diversity is a major factor to be considered, which refers to the degree how different the group members are. There is a sizable literature showing positive effect of the diversity on the achievement of groupwork. It has been argued that people with different demographic backgrounds have qualitatively different life experiences and thus differ in their knowledge, attitudes, and opinions, implying that demographic diversity increases the knowledge pool of the group and as a consequence has positive implications for group cognitive complexity [10][11]. Pelled et al. [12] studied the effects of functional diversity and their hypotheses implied that ability diversity would be positively related to group performance. Jehn et al. [13] found that informational diversity was positively related to group performance and commitment. The diversity in attitudes and values may be associated with positive outcomes (e.g., social integration) or may be unrelated to these outcomes [14]. Although much is still unclear about the effects of diversity and the findings seem inconsistent, this is beyond the discussion in this paper. In this paper we propose a novel grouping algorithm following the below assumption: *Group diversity performs positive influence on the groupwork achievement*.

B. Balanced K-means Algorithm

K-means is one of the simplest yet most popular unsupervised learning algorithms. Given a set of observations $(x_1, x_2, ..., x_n)$, where each observation is a *d*-dimensional real vector, *K*-means clustering aims to partition the n observations into k sets $S = \{S_1, S_2, ..., S_k\}$, so as to minimize the within-cluster sum of squares :

$$\sum_{j=1}^{k} \sum_{x_i \in S_j} \|x_i - \mu_j\| \tag{1}$$

where μ_j is the mean of cluster S_j. A typical *k*-means clustering algorithm consists of the following basic steps:

1) Initialization: Place k points into the space to represent the centroid of each group;
2) Clustering: Assign each object to the group that has the closest centroid;
3) Updating: Update the centroid of the each group when all objects are clustered;
4) Repeating: Repeat step 2 and 3 until termination conditions are satisfied.

The output of the *k*-means clustering is a $n \times k$ matrix with entry a_{ij} which satisfies:

$$a_{ij} = \begin{cases} 1 & x_i \in S_j \\ 0 & \text{otherwise} \end{cases}$$

and $\sum_{j=1}^{k} a_{ij} = 1$, for one object is assigned to one and only one cluster. It is easy to implement and apply this technique even on large data sets. Therefore, the *k*-mean clustering technique has been successfully applied in various area, ranging from statistics, data mining, and information technology [15]. However, in many real world clustering applications, the sizes of each cluster are known in advance due to the application requirements, in other words, $\sum_{i=1}^{n} a_{ij}$ for each n is fixed. More specifically, many applications have size balancing constraints which requires clusters are of approximately the same size, i.e., $|\sum_{i=1}^{n} a_{ip} - \sum_{i=1}^{n} a_{iq}| \leq 1$ for any p and q. In order to address this question, a recent study [4] transformed the problem to a linear optimal problem that find another partition B with balanced size, which minimizes the value of

$$\|AA^T - BB^T\| \tag{2}$$

The value of each entry $b_{ij} \in \{0, 1\}$ in B satisfy:

$$(BB^T)_{ij} = \begin{cases} 1 & i, j \text{ are in the same group in } B \\ 0 & \text{otherwise} \end{cases}$$

Using basic techniques in linear algebra this optimization could be transformed to the following linear programming problem:

$$min(2k - \frac{2}{n} \sum_{j=1}^{k} \frac{1}{\sqrt{a_i b_j}} \sum_{i=1}^{n} a_{ij} b_{ij})$$

This is a typical binary integer linear programming problem which can easily be solved by any existing solver. Therefore, besides the tradition procedure of *K*-means, the balanced *K*-means have one more stage: create size constraints based on the existing knowledge of the data, and then transform the size constrained clustering to a binary integer linear programming problem [4].

III. Methodology

In this section, we first describe the features to be constructed, next present the approach we proposed, at last discuss the algorithm developed.

A. Feature Construction

We utilized the balanced *K*-means algorithm to cluster the students. It is commonly required that machine learning techniques include a vector constructions phase, where the variables are chosen and data is converted to high dimensional vectors, so that the algorithm can be performed and the results can be obtained consequently [16][17]. In this study we emploied three categories diversity features, so-called demographic, functional, and personality diversities, following the typologies in the survey conducted by Knippenberg [3].

TABLE I. FEATURE LIST

Category	Feature	Value Range	Description
Demographic	Gender	$\{0, 1\}$	0: female, 1: male;
	Age	$x \in N$ (Nature number)	the age of a student;
	Ethnic	$\{0, 1\}$	1: White, 0: the rest (Asian, Africa, and others);
	Nationality	$\{0, 1\}$	1: US citizen, 0: non US citizen;
	English Proficiency	$\{1, 2, 3, 4\}$	1: poor, 2: neutral, 3: good, 4: fluent (first language);
	Residence	$\{0, 1\}$	0: US resident, 1: non US resident;
Functional	Quiz1	$\{x \in N \| 0 \leq x \leq 10\}$	the actual grade a student obtained in the first quiz;
	Quiz2	$\{x \in N \| 0 \leq x \leq 10\}$	the actual grade a student obtained in the second quiz;
	Quiz3	$\{x \in N \| 0 \leq x \leq 10\}$	the actual grade a student obtained in the third quiz;
	Monthly1	$\{x \in N \| 0 \leq x \leq 100\}$	the actual grade a student obtained in the first test;
	Monthly2	$\{x \in N \| 0 \leq x \leq 100\}$	the actual grade a student obtained in the second test;
Personality	Interest	$\{1, 2, 3, 4, 5\}$	the feedback of the quesion “I have interests in Mathematics”, 1: strongly disagree, 2: disagree, 3: neutral, 4: agree, 5: strongly agree;
	Attendance	$\{x \in N \| 0 \leq x \leq 8\}$	the times that a student was present when attendance was taken;
	Religion	$\{0, 1\}$	1 : I have religious belief, 0: I have no religious believe;

Demographic diversities. Demographic diversities indicate the distinction amongst readily observable demographic attributes. In this study we employed some attributes including gender, ethnicity, age, nationality, English fluency and residency. A large area of research in behavioral psychology collated evidence in an effort to discover correlations between behavior and gender. How old a person is often determines his/her knowledge and experience. Nationality, residency and English fluency reflect life experiences. Denson and Zhang [18] found that university students benefited from being confronted with different perspectives and interacting with dissimilar others.

Functional diversities. Functional features represent job-related attributes such as differences in educational or functional background. In this study, since all the students share the same education background, and most of them are first year college students, the best information could be used for constructing features might be the performance in the previous quizzes and monthly exams. The grades from 3 quizzes and 2 monthly exams were available by the start of group projects. The quizzes were given in the beginning of a class with a maximum potential score of 10 point, and each quiz took 10 minutes. Regular monthly exams each took 50 minutes with a maximum potential score of 100 points.

Personality diversities. A number of researchers have argued that it is also important to take into account differences that may be neither readily visible nor job-related, such as differences in personality, attitudes, and values. In this study, we constructed three personality features, i.e., religion [19], interest [20] and attendance [21]. The use of interest and attendance to predict students' performance has been explored and the results verified the effectiveness, thus they are considered as crucial personality features. Also, It has been widely reported and commonly accepted that religion has significant effect on the personal behavior.

Table I associates all the features used in this paper. All the features are discrete variables, in which there are some binary variable, whereas, there is huge deviation in ranges for different variables. For example, the dichotomous variables like gender are in the set of (0,1), however, the monthly exams' grades range from 0 to 100. In order to create the vectors for performing *K*-means efficiently, we generate potential *K*-means input vectors by normalizing all the variables/features, which means adjusting values measured on different scales to a notionally common scale (0,1) by simply calculating:

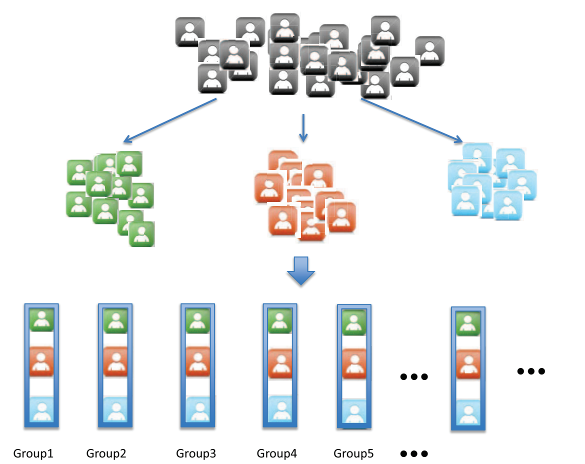

Fig. 1. The procedure of the group organization.

$$z_i = \frac{x_i - min(X)}{max(X) - min(X)} \tag{3}$$

where $X = (x_1, ..., x_n)$ and z_i is i^{th} normalized data.

B. Group Organization

The procedure of grouping has two basic steps:

1) : Apply the balanced *K*-means clustering algorithm to the students' data and get size-balanced clusters. We formalized this step as following:

Given a set of students $(x_1, x_2, ..., x_n)$*, where each student is a d-dimensional real vector, we aim at clustering the* n *students into* k *sets* $S = \{S_1, S_2, ..., S_k\}$*, so as to minimize the within-cluster sum of squares:*

$$\min_S \sum_{j=1}^{k} \sum_{x_i \in S_j} \|x_i - \mu_m\| \tag{4}$$

with the constraints that each cluster is approximately the same size.

2) : Pick up one student from each cluster to form a group. In other words, each group contains k students, who come from the k clusters respectivley. Figure 1 illustrates an example where students are grouped into several 3-member groups. After 3 clusters(color in green, red and blue) are generated in step 1, the students are randomly and uniformly assigned to different groups in the second step.

C. Algorithms

Algorithm 1 A Clustering Based Grouping Algorithm

Require: n studnets $X = \{x_1, ..., x_n\}$
  number of members in one group k
  Balance K-means clustering algorithm L
Ensure: groups G

```
1:  t = n/k {determination of quantity of groups}
2:  for i = 1 to t do
3:     g_i = {} {initialization of each group}
4:  end for
5:  G = {g_1, ..., g_t}
6:  C = {c_1, ..., c_k} = L(X, k) {calling balanced K-means algorithm}
7:  for i = 1 to t do
8:     for j = 1 to k do
9:        if c_j is not empty then
10:          select a random element e from c_j.
11:          c_j = c_j \ e
12:          g_i = g_i ⋃ e
13:       end if
14:    end for
15: end for
```

Algorithm 1 describes the developed algorithm for the proposed grouping model. Besides all the students information in terms of vectors x_i, the number of members in each group k should be given, which also indicates how many clusters are to be constructed. The output is a set of groups G, each group contains k or $k-1$ students. The entire procedure consists of three major stages:

1) **Stage I - Initialization:** In this stage the number of groups to be formed, t, is determined, and each group g_i is initialized (line 1 to line 5);
2) **Stage II - Clustering:** The clusters are obtained by applying the balanced. A set, C is obtained in this stage, each element c_i is a cluster of students K-means algorithm(line 6);
3) **Stage III - Grouping:** The samples in each cluster are randomly and uniformly distributed to the groups (line 7 to line 15);

IV. Experiments

We have conducted a couple of experiments to assess the performance of the proposed grouping approach. We aimed at addressing the following two questions. 1) Is the proposed technique feasible? 2) Is the proposed technique effective? In this section, we first describe the data source and the experimentation setup, and then report result of the analysis, at last discuss the threats to validity.

A. Data Source

We collected data from two mathematics classes in the spring and summer semesters respectively in a university in west Texas. The spring classes (2 sections) totally consisted of 62 students. The students were enrolled in a first-year course, trigonometry, for their bachelor degree of Business. The summer class totally consisted of 21 students, and the participants were undergraduate students major in Engineering at their first or second-year who were taking the class of calculus.

For each class, the students were assigned two separate group projects of the same complexity in which tasks of reading, writing and discussion were involved. Group reports were due in 2 weeks since assignment. It was announced that the grades of both projects would be taken as parts of the final course grades. Students formed groups randomly for the first project, whereas, groups were constructed by the proposed technique for the second one. Group size ranged from 2 to 3 members for both projects. Table II lists the average size for each project in the two classes.

TABLE II. GROUP SIZE

semester	class	project	form	average size
spring	trigonometry	Proj1	random	2.9
spring	trigonometry	Proj2	diverse	2.9
summer	calculus	Proj1	random	2.6
summer	calculus	Proj2	diverse	2.6

The next section descibes how the groups was generated based on the proposed model in this experimental study.

B. Grouping Procedure

To perform students clustering following Section III, we collected three sets of feature information. We obtained demographic and personality features by questionnaire. For the functional features, we simply exported the students' grades from the Grading Center in the Blackboard system[2]. We performed the grouping following 4 steps:

1) copy the raw feature information manually from the feedback of the questionnaire and the record in the Blackboard system;
2) develope a Java utility program to adjust the values of each feature, and build the vectors for each student following the description in section III-A;
3) develope some R scripts to perform the clustering process;[3]
4) develope another Java utility program to form the groups randomly using one-sample-each-cluster strategy.

We carried out the grouping process on a Dell laptop with Windows XP environments. The machine is equipped with a Intel processor (1.66HZ, 2 cores) with 1 GB physics memory. The final group organization information was obtained within 10 seconds after the feature values were ready for use (step 1 excluded which is majorly manual work).

[2] Blackboard is an online teaching systems used in this university

[3] The package "fpc" offers a robust version of K-means based on mediods, which can be invoked by "pam()" function. The package "lpSolve" offers "lp()" function to solve general linear/integer problem.

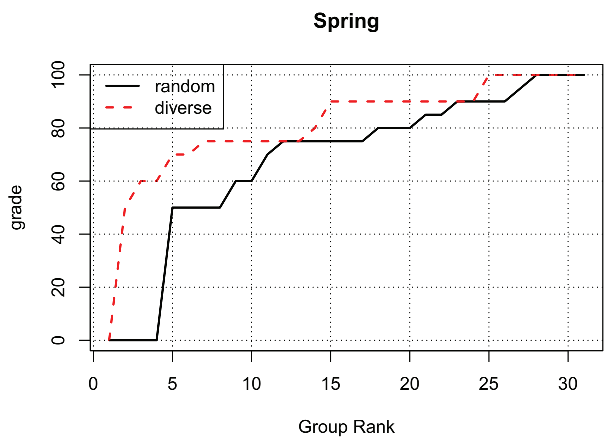

Fig. 2. Groupwork achievement compare in the spring trignometry class (the grades are ranked in ascending order)

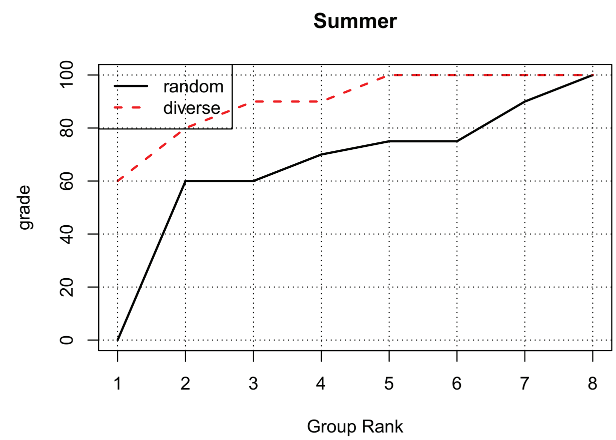

Fig. 3. Groupwork achievement compare in the summer class (the grades are ranked in ascending order)

C. Evaluations

We used the quantity variable, i.e., grades of the projects to measure and compare the groupwork achievement. In order to reduce the threats caused by unintentional bias, each report was graded by three individuals, including one graduate student major in mathematics (teaching assistant) and two instructors with Ph.D. degree in mathematics and computer science respectively. Two of the three graders were unaware of the groups information, in other words, they had no idea that the report they were grading was from a random group or diverse group. The final grade was computed as the average of the three individual scores. We analysed the data from the following two different prospectives.

1) Descriptive Statistics: Descriptive statistics provides simple summaries including quantitative summary statistics like means, standard deviation, and visual, i.e., simple-to-understand graphs. By comparing the means in table III, we observed that the grades of diverse groups were higher than the random groups, and the difference was significant, i.e., around 13 points for the spring class and 25 points for the summer class. On the other hand, the standard deviation of diverse groups was lower than random groups as we expected, due to our one-sample-each-cluster strategy, which made all goups more similar.

We further compared the grades of each group, as Figure 2 and Figure 3 show. The x-axis represents the group's rank accoring to the grades (we ranked all the groups for each project in ascending order), and y-axis represents the score. The figures show that black line(random groups) is underneath the the dotted line(diverse groups) significantly, which indicates the scores of diverse groups are higher than the scores of random groups overall. Therefore, we may conclude that the performance of diverse groups is considerably better.

2) Inference Statistics: Hypotheses test is an essential method in statistical inference [22]. The question of interest is simplified into two competing hypotheses: the null hypothesis, usually denoted by H_0, and the alternative hypothesis, denoted by H_1. More specifically, the null hypothesis H_0 assumes that there is no difference between two groups; whereas, the alternative hypothesis H_1 implies that there is a statistically difference between two sample groups. The probability value (p-value) of a statistical hypothesis test is the probability of falsely rejecting a null hypothesis if it is in fact true (Type I error). Small p-values suggest that the null hypothesis is unlikely to be true. We formulated the null and alternative hypotheses as followings, with the purpose of testing the significance between random grouping and clustering based grouping.

- $H_0 : \mu_{random} = \mu_{diverse}$
- $H_1 : \mu_{random} \neq \mu_{diverse}$

where μ is the mean.

We employed Wilcoxon rank-sum test to perform the significance test. Wilcoxon test is a non-parametric alternative to two-sample *t*-test, which does not rely on the assumption of data complying with any probabilistic distribution [23]. Hence, non-parametric statistical methods are more robust than parametric methods when the underlying distributions are not normal. As Table III shows[4], the p-values for the two classes are smaller than the threshold 0.05, we may come to conclusion that the difference between the two grouping strategies is statistically significant.

TABLE III. WILCOXON TEST RESULT

	Random		**diverse**		
	mean	**stdev**	**mean**	**stdev**	p**-value**
spring	67.9	30.54	81.2	20.19	0.0008426
summer	65.0	31.28	90.0	13.09	0.04758

D. Threats to Validaty

External Threats. We collected data from two mathematics classes, including a spring trigonometry class and a summer calculus class. Different subjects have different charactoristics.

[4]We obtained the value by calling "wilcox.test" function in R

It is of interest to investigate whether the results can be applied to other subjects or not. Also, all the participants are first or second year students who are working for bachelor degree in engineering or business, and further experiments are needed to test students of other majors.

Internal Threats. In both two case studies, project 1 (random groups) was always released and graded one month before project 2 (diverse groups). Internal threats may be introduced by maturation, since the participants are likely to change in terms of their skill and ability such as understanding of the subject and proficiency of reading related materials. Further experiments are needed with different sequence of project 1 and project 2. We extracted the value of features by some self-developed Java utility programs, and performed clustering process and the Wilcoxon test using some R scripts. A random testing methodology was then utilized to test the accuracy of the intermediate values computed.

Construct Threats. We use the quantity variable, i.e., grades to measure the groupwork achievement. In education academia, there are some other evaluation methodologies to assess the groupwork achievement, such as collaborative creativity. Also, in the clustering application scenario, different features usually have various importances to the model. On this topic, the weight of features might be different. For example, gender may play a more critical role compared to nationality. The mono-weighting in this paper may introduce some construct threats.

V. Conclusions and Future Work

In order to implement group diversity automatically and effectively, thus enhancing collaborative learning outcome, in this paper we proposed a novel clustering-based grouping model. We evaluated the proposed technique through two small-scale case studies. Based on the result observed we may come to the conclusion that the proposed technique is practical.

To the best of our knowledge, this is the pioneer work to apply a machine learning technique, clustering, into the groupwork organization theory in education. In order to validate the feasibility and effectiveness of the proposed technique, further experiments are needed based on students on other subjects and majors. In the future we may also search for other techniques [24] to fit in with this scenario, and study the weighting of different features. To sum up, the final aim of our work is to provide a standardized grouping model, which can be used not only in teaching practice, but also can potentially increase the reproducibility of experiments investigating effects of diversity on groupwork achievement for the purpose of research.

Acknowledgment

The authors would like to thank the students who participated the research project. The authors also would like to thank Li Sun and Indri Himawan for their suggestions and helps.

References

[1] P. Dillenbourg, A. Eurelings, and K. Hakkarainen, "Introduction," *What do you mean by collaborative learning*, 1999.

[2] M. M. Chiu, "Group problem-solving processes: Social interactions andindividual actions," *Journal for the theory of social behaviour*, vol. 30, no. 1, pp. 26–49, 2000.

[3] D. Van Knippenberg and M. C. Schippers, "Work group diversity," *Annu. Rev. Psychol.*, vol. 58, pp. 515–541, 2007.

[4] S. Zhu, D. Wang, and T. Li, "Data clustering with size constraints," *Knowledge-Based Systems*, vol. 23, no. 8, pp. 883–889, 2010.

[5] P.-B. Learning, "Speaking of teaching," 2001.

[6] D. H. Dolmans, I. H. Wolfhagen, C. P. Van Der Vleuten, and W. H. Wijnen, "Solving problems with group work in problem-based learning: hold on to the philosophy," *Medical education*, vol. 35, no. 9, pp. 884–889, 2001.

[7] P. McGraw and A. Tidwell, "Teaching group process skills to mba students: A short workshop," *Education+ Training*, vol. 43, no. 3, pp. 162–171, 2001.

[8] S. Maguire and S. Edmondson, "Student evaluation and assessment of group projects," *Journal of Geography in Higher Education*, vol. 25, no. 2, pp. 209–217, 2001.

[9] W. M. Davies, "Groupwork as a form of assessment: common problems and recommended solutions," *Higher Education*, vol. 58, no. 4, pp. 563–584, 2009.

[10] H. Pluut and P. L. Curşeu, "The role of diversity of life experiences in fostering collaborative creativity in demographically diverse student groups," *Thinking Skills and Creativity*, vol. 9, pp. 16–23, 2013.

[11] P. L. Curşeu, S. Schruijer, and S. Boroş, "The effects of groups' variety and disparity on groups' cognitive complexity." *Group Dynamics: Theory, Research, and Practice*, vol. 11, no. 3, p. 187, 2007.

[12] L. H. Pelled, K. M. Eisenhardt, and K. R. Xin, "Exploring the black box: An analysis of work group diversity, conflict and performance," *Administrative science quarterly*, vol. 44, no. 1, pp. 1–28, 1999.

[13] K. A. Jehn and K. Bezrukova, "A field study of group diversity, workgroup context, and performance," *Journal of Organizational Behavior*, vol. 25, no. 6, pp. 703–729, 2004.

[14] D. A. Harrison, K. H. Price, J. H. Gavin, and A. T. Florey, "Time, teams, and task performance: Changing effects of surface-and deep-level diversity on group functioning," *Academy of management journal*, vol. 45, no. 5, pp. 1029–1045, 2002.

[15] Y. Pang, X. Xue, and A. S. Namin, "Identifying effective test cases through k-means clustering for enhancing regression testing," in *Machine Learning and Applications (ICMLA), 2013 12th International Conference on*, vol. 2. IEEE, 2013, pp. 78–83.

[16] X. Xue, Y. Pang, and A. S. Namin, "Trimming test suites with coincidentally correct test cases for enhancing fault localizations," in *Computer Software and Applications Conference (COMPSAC), 2014 IEEE 38th Annual*. IEEE, 2014, pp. 239–244.

[17] ——, "Feature selections for effectively localizing faulty events in gui applications," in *Machine Learning and Applications (ICMLA), 2014 13th International Conference on*. IEEE, 2014.

[18] N. Denson and S. Zhang, "The impact of student experiences with diversity on developing graduate attributes," *Studies in Higher Education*, vol. 35, no. 5, pp. 529–543, 2010.

[19] J. Lodi-Smith and B. W. Roberts, "Social investment and personality: A meta-analysis of the relationship of personality traits to investment in work, family, religion, and volunteerism," *Personality and Social Psychology Review*, vol. 11, no. 1, pp. 68–86, 2007.

[20] M. M. Nauta, "Self-efficacy as a mediator of the relationships between personality factors and career interests," *Journal of Career Assessment*, vol. 12, no. 4, pp. 381–394, 2004.

[21] M. A. Conard, "Aptitude is not enough: How personality and behavior predict academic performance," *Journal of Research in Personality*, vol. 40, no. 3, pp. 339–346, 2006.

[22] M. L. Marx and R. J. Larsen, *Introduction to mathematical statistics and its applications*. Pearson/Prentice Hall, 2006.

[23] S. Siegel, "Nonparametric statistics for the behavioral sciences." 1956.

[24] Y. Pang, A. Wang, X. Xue, and C. F. Martin, "Random walk and linear switching systems," *Communications in Information and Systems*, vol. 12, no. 4, 2012.