

# Dynamic Group Formation as an Approach to Collaborative Learning Support

Ivan Srba and Maria Bielikova, *Senior member, IEEE*

**Abstract**—In the current time of globalization, collaboration among people in virtual environments is becoming an important precondition of success. This trend is reflected also in the educational domain where students collaborate in various short-term groups created repetitively but changing in each round (e.g. in MOOCs). Students in this kind of dynamic groups quite often encounter various difficulties, which are obvious mainly when the students' characteristics do not complement each other. In spite of various group formation methods aimed to solve the group compatibility problem, most of the existing approaches do not consider dynamic groups. We describe (i) a proposal of a novel group formation method based on Group Technology approach, which considers feedback on students' collaboration to improve group formations; (ii) an application of the method as a part of a collaborative platform PopCorm, which provides students in the created groups with a set of real-time collaboration tools; (iii) a long-term experiment in which the groups created by our method achieved significantly better results in comparison with the reference approaches. Our results indicate that considering feedback from students' collaboration can improve the group formation process as the groups created by our method achieved higher collaboration quality with next iterations.

**Index Terms**—Collaborative learning, computer science education, computer uses in education, distance learning



## 1 INTRODUCTION

IN the recent years, collaboration among people became an integral and essential part of the web. Users collaborate and communicate in different kinds of communities and groups across different domains. This trend is present also in educational systems where collaboration is commonly employed not only to share and learn new knowledge but also to develop students' soft skills (e.g. communication skills, self-reflection and self-regulation). Especially, the research area of Computer-Supported Collaborative Learning (CSCL) studies how to effectively link together fast advance in computer science with collaborative learning in small groups [1].

The rising popularity of web-based learning systems caused that many students with different characteristics, skills and aims are supposed to collaborate on common tasks. From one point of view, this diversity has a beneficial effect on creative and successful collaboration. On the other hand, personal differences do not have to be compatible with each other and consequently students' collaboration is not very successful in many cases.

Therefore, collaboration support plays an important role to face this problem. This support is especially substantial during a group's creation process which can significantly influence following collaboration. Thus for effective students' collaboration, we have to solve a challenging task how to successfully identify study groups and help students to find appropriate collaborators [2].

There exist several methods which solve a group formation problem in the educational domain. The significant part of these methods focuses on long-term groups which collaborate on complex tasks during several days or even

weeks. Another part of existing methods is aimed to propose short-term groups but they usually consider only a single assignment of students into groups ignoring following collaboration.

The main subject of our research are *dynamic groups* in which members collaborate on short-term tasks and students are repetitively assigned to groups whose composition differs in each round. This kind of dynamic groups appears especially in online learning systems, in which students learn self-controlled and are not mutually synchronized, such as in Massive Open Online Courses (MOOCs), where:

1. Only limited information about students is available.
2. Rules for successful group creation are unknown or change significantly in time.
3. Groups need to be created ad-hoc and in real time while considering student's actual context and online presence.

All these limitations cause that existing approaches are not very suitable to create dynamic groups. Therefore, the main contribution of this paper is a proposal of a novel method for *dynamic (iterative) formation* of small *short-term* and *virtual study groups* which is supposed to perform better under the given conditions. The proposed method is fundamentally based on its iterative application and on feedback provided by the evaluation of collaboration achieved in the created groups. Following analysis of the existing group formation methods and the requirements of educational context (i.e. collaborative solving of short tasks that exercise primarily new topics), we decided to base the design of our method on the Group Technology approach.

Any such method cannot exist without its application in a real collaborative environment. For this reason, we paid

• I. Srba, and M. Bieliková are with the Faculty of Informatics and Information Technologies, Slovak University of Technology, Ilkovičova 2, 842 16 Bratislava, Slovakia. E-mail: {name.surname}@stuba.sk.

attention to the design and implementation of the collaborative environment too. We introduce a collaborative platform named PopCorm that serves two purposes: 1) as an example how to implement the approach within a learning environment, and 2) as a tool to be used for the method's validation. We are aware of a gap between fast growing collaboration software and its real application in the field of CSCL [3]. PopCorm represents an innovative learning environment with a set of real-time collaboration tools. These tools are based on the latest web technologies and represent an important source of automatically collected feedback to the proposed group formation method.

The paper is organized as follows: we describe group development models with focus on lifecycle phases which are present in dynamic groups in Section 2; Section 3 presents two models of collaboration support: scripting and collaboration management model; our proposed method is introduced in Section 4; we describe its application in the learning environment PopCorm in Section 5 together with a description of performed experiments in Section 6; finally, conclusions are proposed in Section 7.

## 2 GROUP DEVELOPMENT IN ONLINE ENVIRONMENT

The basic concept of CSCL is collaboration which takes place in more or less explicitly defined groups. This collaboration is not performed in one consistent phase. Actually, groups are creating, developing and finally closing. This process can be described as a lifecycle of small groups. Groups' effectiveness and successfulness depends on different circumstances during entire groups' lifecycle [4].

A lot of various models of groups' lifecycle exist. One of the most cited and the most analyzed one is Tuckman's small group development model. In 1965, Tuckman [5] proposed a model with 4 stages of group development: forming, storming, norming and performing. Later, in 1977, Tuckman and Jensen [6] reviewed the original model and added a final stage called adjourning.

Tuckman's model has been already successfully applied to localized long-term study groups (e.g. [7]) but it is not very suitable for groups in online environments which we are interested in. The main reason is that the purpose of stages storming and norming is to build up strong relationships and a common collaboration plan. However, while these attributes play an important role in long-term groups, distributed groups (e.g. those created in various

MOOCs) involve students with more loosely tied relations as well as dynamic groups usually do not solve tasks that require a complex planning. Nevertheless, Tuckman's model becomes the base for many other specialized groups' lifecycle models. One of them is group development model proposed by Daradoumis et al. [4], [8], which was proposed especially for needs of collaborative learning and working in virtual long-term groups.

The main focus of our research is, however, to support short-term virtual groups. These groups exist only for a very short time (usually less than one hour) and thus their lifecycle is simplified in comparison with long-term groups. The phase of productive performing follows immediately after finishing the group formation process. After achieving the group's goal, the short phase of group closing can appear (see Figure 1).

### 2.1 Group Formation

The main goal of the first stage of group development is to solve a problem how to assign students to groups. The traditional approaches to solve this problem are to select students randomly, let students group by themselves or group them manually by a teacher [9]. These approaches, however, have quite substantial disadvantages.

Randomly selected groups can be highly unbalanced what can likely lead to an ineffective composition of groups. Moreover, the random selection ignores any suggestions what a successful group should look like.

The second possibility is to shift the responsibility for group creation to students. Some researches indicate serious problems when the group formation process was managed by students themselves (e.g. [10]). Students tend to create homogenous groups on the basis of existing social relationships or their knowledge level (i.e. good student with other good ones). This trend prevents spreading of knowledge and ideas between students in new social communities. Another problem can be caused by minority students. If they are isolated in groups, this isolation can contribute to more intensive feeling of loneliness which can finally cause their inactivity. Daradoumis et al. [4] in their experiment conclude that 21 out of 138 students in total were not able to find and join any group. By evaluation of questionnaires at the end of the experiment, authors identified the source of this problem. The students did not realize the importance of the group formation process or they became involved in this process very lately.

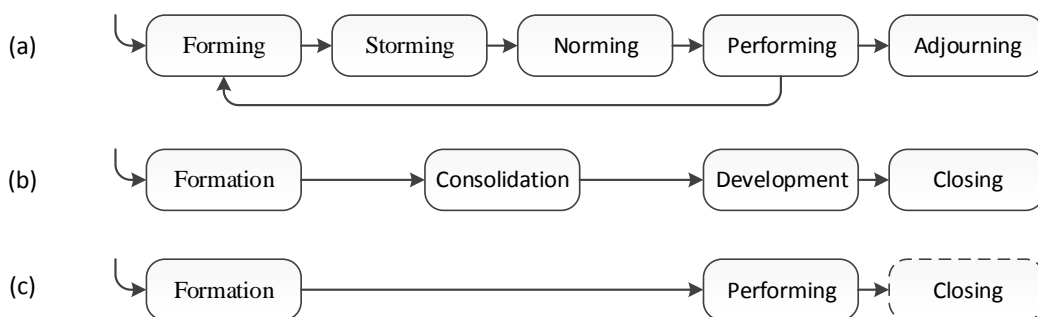


Fig. 1. Comparison of small group development models: (a) the referential Tuckman's model of long-term localized groups [6]; (b) model of long-term virtual groups proposed by Daradoumis et al. [4], [8] and (c) the adaptation of previous two models to short-term virtual groups.

Finally, a teacher can manually assigns students into groups according to information known about students. A teacher can approach this task intuitively and join together those students whose combination he or she believes can lead to active collaboration. This kind of manual group formation can be very difficult and time-consuming [11], especially for a big amount of students or in a case when a teacher does not know students well. In addition, the complexity of this approach increases when we create heterogeneous or mixed groups where the count of all possible group assignments can be really high [12].

In order to create better study groups, automatic computer-supported methods are proposed. Employing computer support in the group creation process can lead to several important advantages. Especially, it is possible to consider a large amount of information even from very different sources. In addition, group creation can be performed very fast and anytime on demand by students or a learning system itself. Last but not least, computer support allows to create anonymous groups in which members do not know their identity.

We identified a big amount of various educational group formation methods. We propose a categorization of different approaches according to the most important attributes of these methods (see Figure 2).

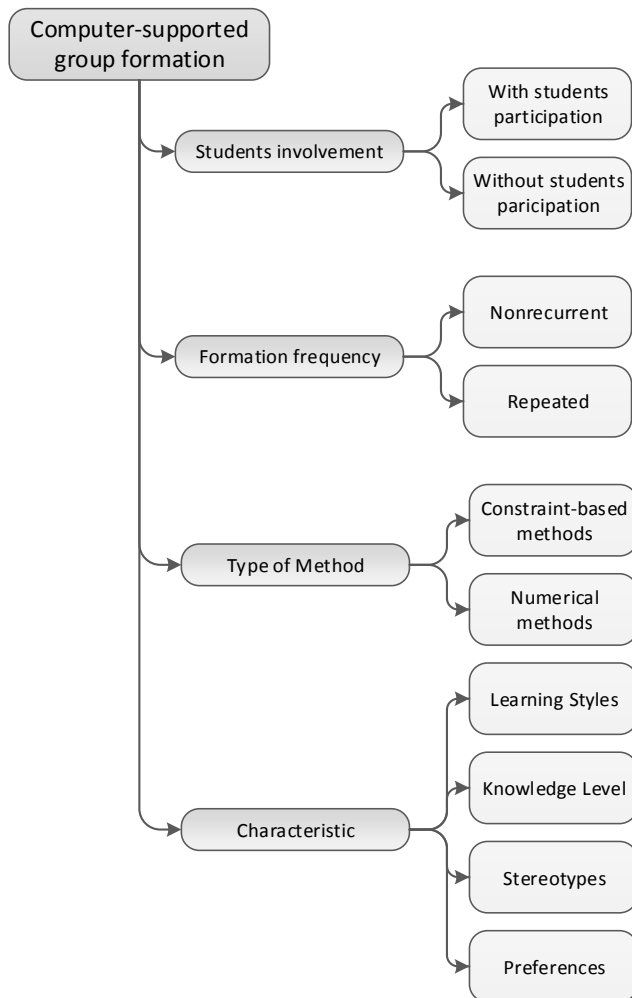


Fig. 2. Categorization of the existing computer-supported approaches to group formation.

**According to students' involvement.** Some methods involve students' participation in the group formation process (e.g. [13]). Students are asked to specify their personal characteristics (e.g. interests or self-evaluated level of knowledge) or preferences (e.g. group or task preferences). Consequently, group formation methods can take advantage of these students' inputs and propose more suitable groups. On the other hand, this approach has several notable disadvantages, such as students may not wish to spend an additional time with filling questionnaires (especially in the case of short-term collaboration). In addition, self-evaluated characteristics can be significantly skewed due to a natural trait of subjective rating.

Another option is that the group formation process can be performed without any active participation of students. In this case, the group formation process usually consists of three steps: initiating a group formation process by a teacher or a learning environment, identifying peer learners who fulfill requirements for participating in the group and negotiating with potential participants [14]. All three steps can be supported by adaptive educational systems.

**According to formation frequency.** Nonrecurring methods for group formation produce a single assignment of students into groups and thus, these methods usually do not consider their following development.

As opposite to this approach, iterative methods suppose that group formation will repeat in several following rounds and, therefore, they can take into consideration feedback from the previous students' assignments.

**According to types of methods.** One of the most used approaches employed in automatic group formation methods is a constraint-based approach in which group formation can be viewed as a constraint satisfactory problem. Students' characteristics together with constraints for group assignments are commonly defined by means of Semantic Web technologies, especially ontologies (e.g. FOAF ontology employed in the approach proposed by [11]). The main disadvantage of these methods is the assumption that a teacher can determine which constraints influence collaboration and make it more effective in all possible situations. However, the current state of the research does not provide a clear answer to this question.

Another type of methods for group formation is numerical methods that do not require exact rules for students' assignments into groups in comparison with constraint-based methods. Student's characteristics are usually represented by an n-dimensional vector where a value in each dimension corresponds to the strength of the particular student's attribute. Two students can be compared by calculating a difference between values of their vectors. This approach allows us to employ any existing technique of clustering for the group formation purpose. Due to data-driven nature of numerical methods, they are successfully applied also in learning environments where education is unstructured and we do not have enough information about students and the educational domain (e.g. [15]).

**According to characteristics.** Last but not least, group formation methods employ different students' characteristics which are used to propose a group composition. Widely used are learning styles (e.g. Felder-Silverman

learning styles model), knowledge levels, personality traits (e.g. personality dimensions according to the well-known NEO-FFI questionnaire) or preferences (e.g. a familiarity with other students) [16].

**Discussion.** Computer-supported group formation methods are able to outperform traditional methods since they are able to consider more extensive amounts of input data. Existing methods consider various sources of students' characteristics, such as questionnaires, social networks, wikis or blogs. However, most of these methods do not consider important feedback from subsequent collaboration (e.g. quality of achieved collaboration). We suppose that this feedback has a potential to significantly improve the group formation process. In addition, most of existing methods assume that it is possible to decide in advance for all cases which aspects make collaboration really effective and successful. However, this information is not well known in the current state of the research.

## 2.2 Group Performing

Group performing is the next important stage of dynamic groups' lifecycle because creating appropriate groups itself is not a guarantee of successful and effective collaboration. Therefore, it is necessary to supply students also with a suitable learning environment which supports collaborative activities by an appropriate computer-based assistance. In order to facilitate collaborative learning, number of tools and functions have been designed and applied in various learning environments so far. They have been classified into 5 categories in [17].

**Appropriate means for dialogue and action.** Tools for dialogue and action represent the most important part of each learning environment. These tools, also termed as groupware, represent the main means for learners in their collaborative activities. They can be divided into two main categories: communication (e.g. an email, a chat) and interaction tools (e.g. a text or a graphical editor). Communication tools are dedicated to supporting activities (e.g. negotiating, decision making, or task management). On the other hand, interaction tools are designated to solve the collaborative task itself and thus their suitability greatly depends on a particular task. Interaction as well as communication tools can support asynchronous or synchronous mode of collaboration.

**Functions for supporting students' self-regulation or guidance.** Besides collaborative communication and interaction tools, it is important to provide students with tools for their self-regulation or guidance. If a learning system presents appropriate visualizations of students' collaborative activities, students' have a possibility to develop their self-regulation and communication skills and thus improve their following collaboration.

**Functions for workspace awareness.** Creating awareness about activities of other users in the shared workspace is essential for achieving effective collaboration [18]. Workspace awareness is defined as "the up-to-the moment understanding of other users' interaction with the shared workspace" [19]. The example of a workspace awareness tool is a participant list with online presence or a position of other participants' text-cursor (e.g. in Google Docs).

**Functions related to community level management.** Besides workspace awareness, it is essential to supply learners also with tools and functions for management of activities and materials produced amongst whole community [17]. This requirement leads to creation of various management tools above learning materials and a community itself (e.g. repositories of learning materials).

**Facilities related to teachers' assistance.** Last but not least, learning environments applied in formal or non-formal learning settings contain facilities which support teacher's or instructor's assistance. The precondition for providing assistance to learners is an overview about activities currently performed in a learning environment. A teacher can benefit from individual, collaborative and even comparative information based on analysis of all interactions [17]. Quite significant part of the research in Technology Enhanced Learning (TEL) is concerned with learning analytics aiming to provide teachers as well as students with appropriate information about their collaboration. Results of learning analytics are usually presented by different kinds of visualizations or dashboards.

**Discussion.** Collaboration and its effectiveness substantially depends on available tools and functions. However, despite rapid development of collaborative tools outside the educational domain, learning systems only very slowly adapt modern techniques (e.g. real-time collaboration).

## 2.3 Group Closing

The lifecycle of short-term virtual groups can sometimes consist of the third phase which is group closing. It is a very short stage during which members of a particular group have a possibility to review the achieved solution and collectively confirm the completion of the task being solved. During group closing, students usually use for negotiation the same standard communication tools, which are employed during the group performing phase.

# 3 COLLABORATIVE LEARNING SUPPORT

Despite many studies (e.g. [20]), which confirm that collaborative learning correlates with a wide range of positive outcomes (e.g. improved learning, increased productivity, higher motivation), collaborative learning does not work automatically for all learners [21]. This is especially true for short-term groups in which members have not cooperated before and their individual goals are predominant [22]. Therefore, it is important to provide students with educational support during whole groups' lifecycle which is referred as to scaffold collaboration.

Particular methods to provide educational support can be based on various underlying collaboration scaffolding models. In general, there are two main complementary approaches [23]: by structuring the collaborative process (commonly by scripting) or by regulating (as widely used collaboration management model does).

## 3.1 Scripting

Methods based on scripting are mostly employed in the CSCL domain during learning process where unconstrained collaboration does not lead to expected results. A



collaboration script is a predefined set of instructions prescribing particular phases of collaboration, e.g. how students should form groups, how they should interact and how they should solve an assigned problem [23]. Scripting can occur at different levels of granularity [24]:

1. Macro-scripts are high-level models which describe a sequence of activities performed by users who play usually different roles.
2. Micro-scripts are dialogue models which are directly embedded in collaborative environments.

Scripting was confirmed as a promising approach of scaffolding collaboration which results in improved learning [25]. However, at the same time, scripting is criticized for restricting users' freedom and independence. This phenomenon is called over-scripting [23]. Adaptive scripting methods were proposed to deal with this problem by definition of elements that can be easily adapted but without reducing the added value of the collaborative process [25]. These elements are called intrinsic. On the other side, extrinsic elements are those which cannot be adapted in any way (e.g. due to technological or pedagogical restrictions).

### 3.2 Collaboration Management Model

The concept of collaboration management model was introduced by [26] and was confirmed as a successful way how to scaffold collaborative learning [27]. In comparison to scripting, it is based on decisions made in run-time rather before collaboration begins. Collaboration management model refers to a simple process of continuous comparing a current state of collaboration with a desired state. This process consists of four phases (see Figure 3):

1. The data collection phase involves observing students' interaction. User activities are recorded as logs which are stored for later processing.
2. In the second phase, obtained log records are processed to derive high-level variables called indicators. Afterwards, the current state of interaction is represented by a model of interaction, which consists of a set of indicators. These indicators represent any attribute of the collaboration process, such as an average time delay between activities.
3. In the next step, the acquired model of the current interaction is internally compared with a model representing the desired state of interaction.

4. And finally, if there are any discrepancies between the current and the ideal model of interaction, the system can advise or recommend users how to suppress this undesirable difference.

Moreover, authors proposed a categorization of collaborative systems according to the number of phases of collaboration management which are performed:

1. Mirroring tools which only collect raw interaction data (phase 1);
2. Metacognitive tools which derive a model of interaction (phase 2) and optionally compare it with an ideal model (phase 3);
3. Guiding systems which advice how to improve collaboration (phase 4).

### 3.3 Towards Dynamic Group Formation

As the lifecycle of dynamic groups omits the warm-up phases (i.e. storming and norming phases as described in Tuckman's model or the consolidation phase introduced in the model proposed by Daradoumis et al.), the composition of dynamic groups has a very strong influence on following collaboration. Therefore our main intent is to scaffold students' collaboration primarily by means of the group formation phase. We combine the concepts of both scaffold models, scripting as well as collaboration management model, to provide dynamic groups with a complex collaborative environment which supports learning by means of adaptive group formation. The core of this environment is a group formation method which we describe in the following section.

## 4 DYNAMIC GROUP FORMATION

According to the state of the art in computer-supported group formation techniques, we identified several drawbacks of the existing methods which cause that these methods are not very suitable to create dynamic groups that are created on demand while various domain-specific restrictions have to be considered (e.g. to involve only those students who are currently online). The existing methods commonly do not consider results achieved by the created groups; rely on well-specified rules how to create successful and effective groups; and are too static to be employed in online and dynamic environments. Moreover, they mostly produce only a single partition of all students into

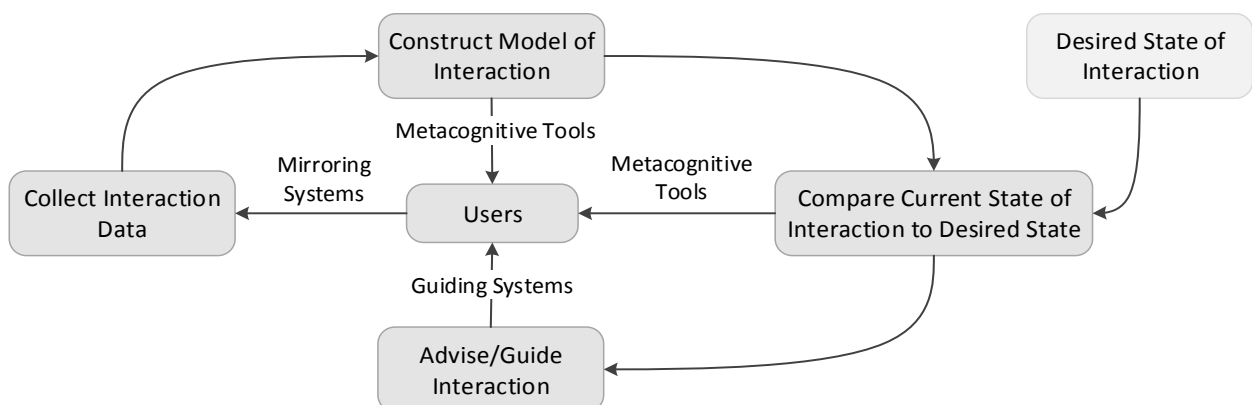


Fig. 3. Collaboration management cycle as proposed in [26].

groups which is not suitable to create dynamic groups. Following this motivation, we introduce a method [28] which is designed to meet three requirements:

1. Feedback from the previous collaboration activities needs to be taken into account.
2. The method has to continuously learn how successful groups should be created.
3. The method has to create groups in real-time and students have to be served fast.

The last requirement means that as soon as students finish solving a task in a group, the achieved results must be reflected immediately in the following group proposals.

With the reference on the proposed categorization of computer-supported group formation methods (Section 2.1), the proposed method does not involve students in the group formation process, belongs to the group of iterative and numerical approaches, and is independent on particular characteristics.

#### 4.1 Problem Formalization

The main goal of our method is to propose study groups by finding compatible learners on the basis of their individual characteristic. We consider learners' characteristics as *compatible* when their combination leads to positive outcomes (e.g. high and low level of knowledge about a particular domain topic).

Input data to the proposed method are composed of two matrices: 1) a matrix of characteristics' compatibility and 2) a matrix of assignments of the characteristics to students (see Figure 4).

The matrix of characteristics' compatibility is defined as follows. Let  $C$  be the set of all characteristics  $C = \{c_j\}$ ,  $j = 1, 2, \dots, n$ . Every characteristic is represented by an  $n$ -dimensional vector  $c_j = (c_j^1, c_j^2, \dots, c_j^n)$ , where:

$$c_j^i = \begin{cases} 1 & \text{if } c_j \text{ should be combined with } c_i \\ 0 & \text{if } c_j \text{ should not be combined with } c_i \end{cases}$$

The matrix of assignments of characteristics to students is defined as follows. Let  $L$  be the set of all learners  $L = \{l_k\}$ ,  $k = 1, 2, \dots, m$ . Every learner is represented by an  $n$ -dimensional vector  $l_k = (l_k^1, l_k^2, \dots, l_k^n)$ , where:

$$l_k^i = \begin{cases} 1 & \text{if characteristic } c_i \text{ is typical for learner } l_k \\ 0 & \text{if characteristic } c_i \text{ is not typical for learner } l_k \end{cases}$$

The expected output data from the method are clusters of compatible students. A particular study group can be created with any combination of students from the same cluster. The specific combination and the number of students depends on domain-specific preferences. It means that besides characteristics' compatibility, we can take into consideration any preferences and suggestions how groups should be formed – it is possible to utilize: 1) previous group assignments (e.g. combine only students who have not collaborated together so far or on the other side preserve group stability to avoid too many group switches); 2) knowledge which is necessary to solve the assigned task (e.g. create a group where all required roles are present); or 3) technology-specific preferences (e.g. involve only those students who are currently online).

After the group finishes solving of the assigned task, collaboration and the achieved results are evaluated. Each combination between characteristics present in the group is strengthened according to the achieved evaluation. It means that the better students' collaboration was the more compatible the characteristics are.

In addition, some students' collaborative characteristics (e.g. those that describe collaborative behavior) can be automatically derived from group interaction, too. Therefore, it is possible to consider student's activities to update the matrix of assignments of characteristics to students. The precondition is that there is a technique how to automatically analyze students' collaboration and identify expressions of these characteristics (e.g. by means of learning analytics or sentence openers as we propose in Section 5.1).

The input matrices may be filled in various ways, such as by questionnaires, external sources (e.g. academic information systems, social networks) or existing user models. However, since the method is fundamentally driven by feedback from group interaction, its true power is that the input matrices do not have to be known at the beginning of collaboration at all. The cold start problem can be solved by the random composition of the first groups. Consequently, both matrices are continuously learned and improved by means of returned feedback. Moreover, this approach enables the required adaptation to changing conditions under which groups achieve positive results.

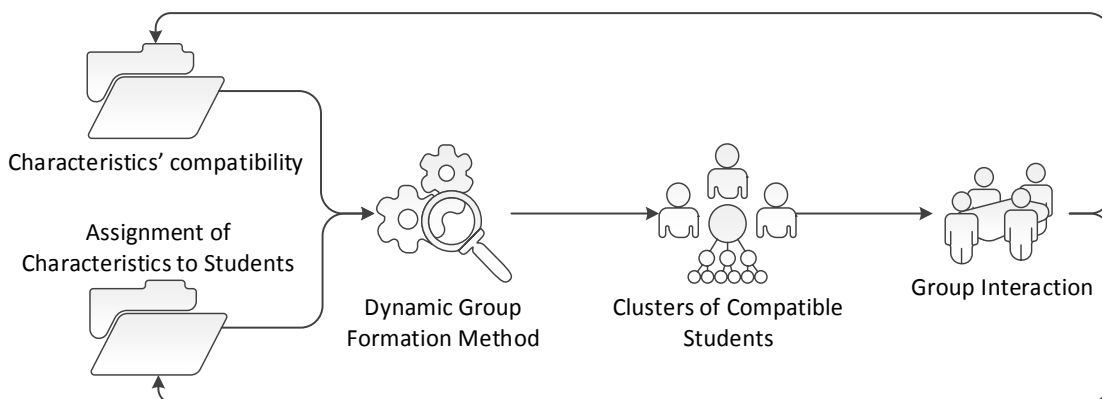


Fig. 4. Schema of the proposed method. An input to the method consists of two matrices: a matrix of assignments of characteristics to students and a matrix of characteristics' compatibility. An output from the method consist of clusters of compatible students.

## 4.2 Group Technology Approach

Recently, many methods and techniques developed for various domains were applied to group formation in the educational domain, e.g. genetic algorithms [29], particle swarm optimization [30] or ontologies [11]. On the basis of the stated requirements on the proposed method, we decided to employ Group Technology approach.

According to Selim, et al. [31] Group Technology (GT) is an approach to manufacturing and engineering management that helps manage diversity by capitalizing on underlying similarities in products and activities. The main task in GT approach is so-called Cellular Manufacturing problem, which is inspired by the design of optimal distribution of machines which cooperate on production of a set of parts' families. It is necessary to identify families of similar parts and machines to solve the problem of optimal distribution of cooperating machines. This process is called cell formation. In other words, groups of machines should be located in the close proximity in order to produce a particular family of similar parts and thus minimize the production and transfer time [32].

Group Technology approach seems to solve a similar problem as we have. Analogy between domain entities can be easily found. It is possible to replace a machine with a student, a part with a characteristic, an assignment of parts to the machine with an assignment of characteristics to the student, and a family of similar parts with a set of compatible characteristics. Moreover, we can find this analogy also in goals; instead of optimizing a machine production we need to optimize a group composition.

Several works employing Group Technology approach in the CSCL domain exist. Pollalis, et al. [33] proposed a method for learning objects' recommendation to study groups according to students' knowledge of domain terms. Two input matrices were used. The first one represented student's knowledge; the second one represented similarity or mutual dependency of domain terms which was derived from common occurrence in the same learning object. The output was clusters of students and learning objects which were suitable for these students to learn.

Similar approach is described in [32] and [34]. The main goal of this research was to identify sets of students which use similar strategies to solve mathematical exercises. Similarly to the previous work, two matrices were calculated: the dynamic matrix representing an assignment of strategies to students and the static matrix representing mutual similarity of strategies. The output was clusters of students and assigned groups of strategies. The identified clusters are used to assign a new task to a particular group of students according to strategies familiar to the members of the group and suitable to solve this task as well.

As opposed to the previous two works, authors in [35] considered only one matrix. This matrix represented teachers and subjects they teach. A hybrid grouping genetic algorithm was used to identify groups of similar subjects.

The described works document achieving interesting results in the experiments with the methods based on GT approach in the educational domain. It supports that GT can be applied in the educational domain and support effective education in spite of its technological background.

## 4.3 Group Formation Based on Group Technology

Our dynamic group formation process represents a standard cell formation problem as described in Section 4.2. Several approaches to solve the problem of cell formation are described in [31]. The most appropriate for our goal are procedures based on cluster analysis, especially array-based clustering techniques. In the proposed method, the calculation is performed in several steps:

1. Calculation of vectors' comparison values.
2. Calculation of similarity and relevance coefficients.
3. Creation Group Compatibility Matrix.
4. Clustering on Group Compatibility Matrix.

**Calculation of vectors' comparison values.** First of all, three values are defined for each learner vector  $l_k \in L$  and characteristics vector  $c_j \in C$ . These values are calculated by comparison of these vectors as follows:

1. Value  $a$  is the number of characteristics contained in both vectors.
2. Value  $b$  is the number of characteristics which learner  $l_k$  has but are not compatible with the characteristic  $c_j$ .
3. Value  $c$  is the number of characteristics which the particular learner  $l_k$  does not have but are compatible with the characteristic  $c_j$ .

**Calculation of similarity and relevance coefficients.** Similarity (SC) and relevance coefficients (RC) are defined with these three values. Similarity coefficient is actually well-known Jaccard coefficient and represents how the user is related to the characteristic. On the other hand, relevance coefficient expresses how well the characteristic is compatible with characteristics which the user already has.

$$SC(l_k, c_j) = \frac{a}{a + b + c}$$

$$RC(l_k, c_j) = \frac{a}{a + b}$$

**Creation Group Compatibility Matrix.** By means of similarity and relevance coefficients, Group Compatibility Matrix,  $GCM = (a_{ij})$ ,  $i \in [1, n]$ ,  $j \in [1, m]$ , is calculated as:

$$a_{ij} = \begin{cases} 1 & \text{if } SC \geq \theta^{SC} \text{ and } RC \geq \theta^{RC} \\ 0 & \text{else} \end{cases}$$

Values  $\theta^{SC}, \theta^{RC} \in (0,1)$  represent minimal thresholds for similarity and relevance coefficient. Algorithm starts with thresholds set to 1 and continuously decreases them until a valid GCM matrix is found. The GCM matrix is valid as soon as each student has at least one assigned characteristic.

**Clustering on Group Compatibility Matrix.** Finally, it is necessary to perform clustering on the GCM matrix with any array-based clustering algorithm. For our purpose Modified Rank Order Clustering (MODROC) was used. Output data from our method is the GCM matrix in which the clusters of students and the characteristics are concentrated along the main diagonal (see Table 1). An assignment of a student to a cluster means that the student has these characteristics or these characteristics are compatible with characteristics which are typical for the student. Identified clusters of students represent the required output of the proposed method.

**TABLE 1**  
AN EXAMPLE OF CLUSTERED GCM MATRIX ACQUIRED IN THE QUALITATIVE EXPERIMENT

Collaborative characteristic	L1	L2	L3	L4	L5	L5
Warn of mistake	1	1	0	0	0	0
Accept warn of mistake	1	1	0	0	0	0
Write comment	1	0	0	0	0	0
Write general message	0	0	1	0	0	0
Ask for explanation	0	0	0	1	1	1
Give explanation	0	0	0	1	1	1
Propose action	0	0	0	1	1	1
Accept action	0	0	0	1	1	1
Write praise	0	0	0	1	0	0

#### 4.4 Extension of Group Formation Method

It is possible to extend the proposed method by consideration of several categories of characteristics, e.g. the first category can represent demographic information about students (e.g. age, gender) and the second one can represent collaborative characteristics (e.g. argumentation and reaching consensus).

The separate matrix of characteristics' compatibility is used for each category of relevant characteristics and our method is applied on each matrix individually. As the result, several GCM matrices are obtained and thus each student is assigned to as many clusters of students and characteristics as the number of characteristics' categories is. It means that we can combine these clusters to create even more appropriate groups. Similarly as in the basic version of the method, specific way how to combine clusters depends on particular categories of characteristics and domain specific requirements. An example of students' assignment to several clusters of students and characteristics is displayed in Table 2. If we receive a request to assign the student  $l_1$  to a new group, we can for example:

1. Focus on student's actual context in the learning system and use those clusters of students and characteristics which are relevant to his or her actual context, i.e. if the student  $l_1$  is reading a learning object about design patterns, we can use a combination of specialization and knowledge of programming languages. We will achieve that students in the created group will be able to talk about applying design patterns in the well-known domain (web applications or DB systems) and in the familiar programming language (Ruby).
2. Another possibility is to create a new group without focus on the particular category of related categories. We can consider only those students who have common all categories of characteristics with student  $l_1$  ( $l_6$ ) or at least two categories ( $l_3, l_6, l_7$ ).

## 5 COLLABORATIVE ENVIRONMENT FOR DYNAMIC GROUP FORMATION

As we stated before, the proposed method is the core part

**TABLE 2**  
EXAMPLE OF AN ASSIGNMENT OF STUDENT  $l_1$  TO SEVERAL CLUSTERS OF STUDENTS AND CHARACTERISTICS

Category of related characteristics	Cluster of characteristics	Cluster of students
Specialization	Web applications, DB systems	$l_1, l_4, l_6$
Knowledge of programming languages	High knowledge of Ruby, medium knowledge of PHP	$l_1, l_3, l_6, l_7$
Collaborative behavior	Active, communicative	$l_1, l_2, l_3, l_6, l_7$

of our collaborative environment. The design of collaborative environment is, however, substantially domain specific. We decided to apply our method in the formal learning settings in which students assigned into dynamic groups solve short-term practical tasks that supplement a one-term class, which is held at a *university*. More specifically, we situate the collaborative environment into context of a class dedicated to the basics of *software engineering*. We implemented this kind of collaborative environment as a collaborative platform named *PopCorm (Popular Collaborative Platform)*.

### 5.1 Guiding System based on Micro-scripts

According to the analyses of the most common approaches to scaffolding collaboration (see Section 3), we designed collaborative platform PopCorm as a guiding system which capitalizes on the positive effects of micro-scripting.

**Collect interaction data.** Current information technologies are able to capture students' overall interaction quite precisely. However, it is possible to automatically capture interaction only on a very low-level (e.g. a plain text of a message sent in a chat). To describe interaction better, there are several approaches which include natural language processing (e.g. sentiment analyses) or data mining (e.g. categorization or sequential pattern mining). However, it is not a trivial task to apply these approaches in collaborative learning environments because despite their significant improvement in the recent years, they are still limited in understanding and interpreting communication [36].

Therefore, we decided to employ another solution which is to structure the collaborative interface by micro-scripting approaches, and more specifically by sentence openers. Sentence openers refer to a communication interface in which users select a beginning of a sentence (e.g. "I suggest to...") and complete the sentence with the rest of the message they would like to communicate with others.

This kind of micro-scripting techniques allow us to automatically describe interaction also on a high level. Moreover, groups, in which members communicate via a structured interface, show more intensive orientation on finding the solution in comparison with groups in which members communicate via an unstructured interface [37]. In addition, students by themselves tend to use a structured discussion (the experimental study performed in [37] shows that the structured messages represented about 58% of all



sent messages). On the other hand, strictly structured communication interface can negatively influence collaboration, especially in cases where students need to communicate in a way which is not adequately supported by the provided interface (so called problem of over-scripting).

In PopCorm, we applied the sentence openers to structure communication among students by means of 16 different types of messages selected according to McManus and Aiken's taxonomy of Collaborative Skill Network [38]. This taxonomy defines a set of sentence openers that correspond to conversation skills commonly used during collaborative learning and problem solving, such as propose a better solution, accept a proposal, ask for an explanation or provide an explanation. To solve the problem of too strictly structured interface (over-scripting), we decided to include also 2 additional types of messages: a general message and a comment, which can be used when students cannot assign their message to any of the predefined types.

**Construct model of interaction.** Consequently, the captured interaction is used to derive the high-level variables describing the collaboration process. There are several models how to evaluate technology-mediated collaborative learning. Authors in [39] consider a multidimensional model proposed by [40] as the most representative one and refined this model to include 7 indicators of collaboration quality which we adopted in the design of the collaborative platform:

1. Sustaining mutual understanding ( $I_1$ ).
2. Information exchanges for problem solving ( $I_2$ ).
3. Argumentation and reaching consensus ( $I_3$ ).
4. Task and time management ( $I_4$ ).
5. Sustaining commitment ( $I_5$ ).
6. Shared task alignment ( $I_6$ ).
7. Fluidity of collaboration ( $I_7$ ).

Each of these indicators is automatically calculated on the basis of recorded activities in the proposed structured interface. The indicators  $I_{1-5}$  are calculated as a proportion of positive activities (i.e. those that positively contribute to

the particular dimension) in all activities (i.e. a sum of activities with a positive and negative influence). The indicators  $I_{6-7}$  are calculated as an equality in distribution of activities among members or in time respectively. In addition, a teacher can manually add the eighth indicator representing a quality of the created solution itself ( $I_8$ ). All indicators are represented by a value in the interval  $(0, 1)$ .

An overall evaluation of collaboration quality ( $OE$ ) is calculated as a weighted arithmetical average (with strengthen teacher's evaluation of the achieved solution) from all eight indicators as follows:

$$OE = \frac{\sum_{i=1}^7 I_i + 3 I_8}{10}$$

**Compare current state of interaction to desired state.** The desired state of collaboration is represented by the indicators which values are equal to 1s. In other words, the desired state of evaluation is achieved when the overall evaluation reach value 1.

**Advise/guide interaction.** After finishing solving the tasks, the collaborative platform PopCorm provides students with visualizations of all 8 indicators. In addition, it advises students how to collaborate better and more effectively if some of these indicators reach only very low values in comparison with the ideal state of collaboration.

## 5.2 Collaborative Tools and Functions

To provide students with an attractive system, we based PopCorm's implementation on several concepts well-known from content creation tools (e.g. Google Docs), such as collaboration in real-time or a timeline of content evolution (for further information about PopCorm see [41]). Moreover, PopCorm provides all functions and tools which are essential for effective and successful group performance (see Section 2.2). According to the requirements of the selected course, we have recognized the need to design one communication tool: a semi-structured discussion; and three interaction collaborative tools: a text editor, a graphical editor and a categorizer (see Figure 5).

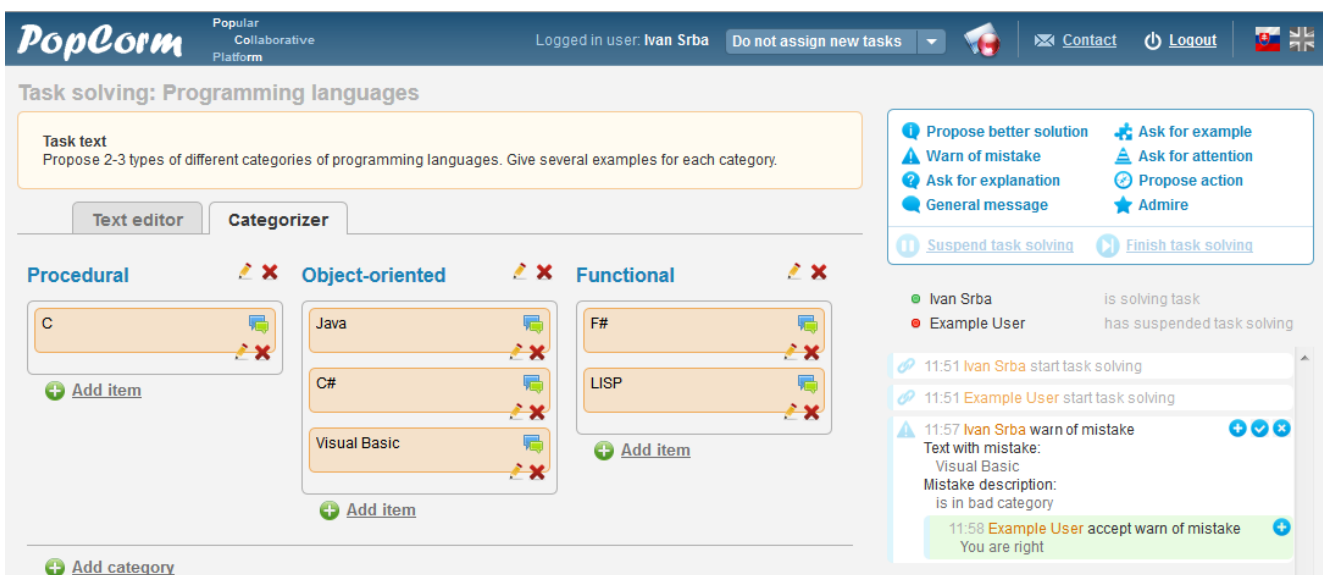


Fig. 5. Screenshot from the collaborative platform PopCorm; the categorizer tool is displayed on the left side. The sentence openers are available in the upper right corner. Below the sentence openers, there is the online presence with further information about student's current activity. In the lower right corner, the history of communication is placed.

The *semi-structured discussion* represents a generic communication tool independent of the particular type of the task at hand. As the discussion represents a main tool for communication, it is based on the sentence openers approach (previously described in Section 5.1) and provides 18 types of messages, which we decided to use to structure the communication among students.

The *text editor* as an interaction tool is suitable for collaborative writing of free text. It provides functionality for parallel editing of written text by several users at the same time together with the conflict resolution in the case when two users edit the same part of the text simultaneously. In addition, the text editor promotes students' authorship by colored text highlighting (each group member has assigned his or her unique color). We based the implementation of the text editor on the open source online editor Etherpad (<http://etherpad.org>).

The *graphical editor* provides the opportunity to collaborate visually by draft drawing, especially by designing UML diagrams. Its functionality covers drawing vector shapes, importing raster images and adding text notes. As well as the text editor, also the graphical editor was designed to support content synchronization in real time. It means that students can collaborate on the same drawing without any restrictions. In the implementation of the graphical editor, we proceeded from the drawing editor SVG-edit (<https://code.google.com/p/svg-edit>).

Last but not least, the *categorizer* is a special tool developed for solving different types of tasks that result in one or more lists (categories). The categorizer allows learners to create categories or items, move items from one category to another and reorder items in categories with a standard drag-and-drop technique.

Other tools provided by PopCorm include user online presence, students' profiles, administration, and student or class-wide statistics dedicated to teachers.

## 6 EXPERIMENTS

The collaborative platform PopCorm became the main means to evaluate the proposed method. We evaluated our method and the collaboration platform in two phases. Firstly, we performed a qualitative experiment with several selected participants. Secondly, a quantitative experiment was conducted to evaluate the performance of dynamic groups on a wider audience of students.

### 6.1 Experimental Setup

During both experiments, PopCorm was integrated with an educational system ALEF [42]. ALEF is dedicated primarily for individual learning and indirect collaboration. Indirect collaboration in ALEF includes text annotations (i.e. text highlighting, tagging, error reporting), supplementing learning materials with external sources or collective evaluation of test answers [43]. Therefore, students were able to use ALEF as a valuable source of information while solving collaborative tasks in PopCorm.

**Characteristics.** In the experiments, we used two categories of characteristics to illustrate the universal design of our method: collaborative skills and personality traits.

*Collaborative skills* refer to collaborative learning conversation skills defined in McManus and Aiken's Collaboration Skill Taxonomy. They can be automatically assigned to students according to the most used messages in the semi-structured discussion. The matrix of assignments of these characteristics to students was unknown at the beginning of the experiment and it was continuously learned. Similarly, the matrix of collaborative skills' compatibility was continuously updated according to the overall evaluation achieved in the created groups.

As *personality traits* we used assignments of students into 5 domains of adult personality: extraversion, neuroticism, conscientiousness, openness to experience and agreeableness. To obtain these assignments, students, who participated at the experiments, filled out NEO Five Factor Inventory (NEO-FFI) questionnaires which were evaluated by a team of psychologists. Some studies (e.g. [44]) suggest which combinations of NEO-FFI personality dimensions lead to better study results. However, results of these studies are not very representative and thus we decided to derive personality traits' compatibility from feedback provided to the method similarly as in the case of collaborative skills.

**Collaborative tasks** were defined in a way that enables active participation of every group member. In our experiments, the tasks were created by a teacher. Alternatively, tasks can be created by students themselves. We used 70 collaborative tasks according to seven different types of tasks which are suitable for the domain of software engineering, nevertheless they can be easily used also in a number of different domains:

1. Group discussions about any general problem, e.g. discuss under which circumstances it is suitable to develop software with agile methods.
2. Explanations of domain relevant terms, e.g. explain what a composition and an aggregation means in data modeling.
3. Proposals to some well-defined problems, e.g. propose a state diagram of a bug report in issue tracking systems.
4. Listings of particular items, e.g. list at least five most important software attributes of an application for electronic banking.
5. Comparisons of two entities, e.g. compare COTS and MOTS applications.
6. Advantages/Disadvantages of an entity, e.g. give advantages and disadvantages of use case points.
7. Pros/Cons of two entities, e.g. compare two models of software lifecycle. The first student defends a waterfall model, the second one defends an iterative model.

### 6.2 Qualitative Experiment

The purpose of the qualitative experiment was to evaluate attributes of the proposed method; namely, how well the proposed method is able to identify the clusters of compatible students and consider feedback from the created groups to learn both input matrices. In addition, the experiment was also an opportunity to get valuable comments on the implementation of the collaboration platform. Five

TABLE 3  
STATISTICS OF THE RESULTS ACHIEVED IN THE QUANTITATIVE EXPERIMENT

Metric	Value	Additional notes
Number of students	110	Students attended the 2 <sup>nd</sup> year of the bachelor study (the average age was 21 years).
Number of groups	254	Additional 35 groups were created but students were not able to start collaboration.
Number of activities	3 763	Each activity corresponds to one sent message in the semi-structured discussion.

participants in total took part in the qualitative experiment and solved 12 tasks. The proposed method was able to identify three clusters of students and collaborative characteristics at the end of the experiment (see Table 1).

### 6.3 Quantitative Experiment

The second phase of our evaluation consisted of the quantitative experiment which was realized during the summer term 2011/2012 as a part of the bachelor degree course Principles of Software Engineering at the study programme Informatics, Slovak University of Technology in Bratislava.

Detailed statistics of the quantitative experiment are provided in Table 3. The experiment lasted approximately 2 months. During this time, 110 students in total voluntarily participated in 254 created groups and 3 763 activities were recorded. We found out that students were able to solve the prepared tasks in 11 minutes in average.

For both categories of characteristics, the performance of groups created using our method was compared with a reference method and a traditional approach (randomly created groups). As the reference method, we decided to use a numerical method based on k-means clustering. For purpose of the clustering, each student was represented by a vector of his or her assignments of characteristics.

Particular groups were created from online students either randomly or from the clusters that were derived from collaborative skills or personality traits as well as created by the proposed or reference method. A balanced mechanism was used to employ all combinations of methods and categories of characteristics equally. The group size was restricted to 2 or 3 members (triads were preferred when the sufficient number of students from the same cluster of compatible students were online at the same time).

Consequently, the most suitable task was assigned to the created group. Tasks that 1) have not been solved be-

fore by any member of the group; and 2) practice just lectured topics were preferred. As soon as students finished solving a task in one group, they were asked whether they wanted to continue in collaborative learning in another group (with a different composition).

For each type of groups, we compared the automatic evaluation of collaboration quality (by 7 high-level indicators described in Section 5.1), the teachers' manual evaluation of the created solution (the eighth high-level indicator) and the overall evaluation. The experiment was double-blinded so teachers as well as students were not informed about the method used to create the particular group.

As the results show (see Table 4), groups created by our method achieved more effective and successful collaboration in comparison with other two approaches in all three kinds of evaluation. In addition, the groups created by the proposed method achieved better evaluation for both categories of characteristics that have been used as inputs for the evaluated methods.

In general, the achieved evaluation of groups can be rated as satisfying in spite of the fact that the groups did not achieve the state of ideal collaboration (the overall evaluation and the partial evaluations should theoretically reach values 1, however, this value represents collaboration which can be hardly achieved in practice, e.g. all members would have to send exactly the same number of messages in the discussion).

We evaluated also students' subjective perception of collaboration. Students were asked to provide explicit feedback at a 5-point scale after finishing collaboration (1 means poor collaboration and 5 means excellent collaboration). When collaborative skills were considered, groups created by the proposed method achieved a notably higher feedback evaluation in comparison with other types of groups. For personality traits, higher feedback was achieved in the groups created by the reference method.

TABLE 4  
COMPARISON OF THE RESULTS ACHIEVED BY GROUPS CREATED BY THE EXAMINED GROUP FORMATION METHODS

Characteristics Used as input to method	Group Formation Method	Overall evaluation Interval $\langle 0, 1 \rangle$	Collaboration $I_1-I_7$	Solution $I_8$	Feedback Interval $\langle 1, 5 \rangle$
Collaborative skills	Proposed (Group Technology)	<b>0.451 ± 0.125</b>	<b>0.36 ± 0.12</b>	<b>0.69 ± 0.23</b>	<b>3.96 ± 1.50</b>
	Reference (k-means)	0.380 ± 0.114	0.29 ± 0.12	0.61 ± 0.20	3.40 ± 1.65
	Traditional (Random selection)	0.404 ± 0.122	0.32 ± 0.13	0.62 ± 0.23	3.18 ± 1.95
Personality traits	Proposed (Group Technology)	<b>0.456 ± 0.125</b>	<b>0.36 ± 0.13</b>	<b>0.72 ± 0.21</b>	3.63 ± 1.58
	Reference (k-means)	0.322 ± 0.117	0.29 ± 0.14	0.40 ± 0.25	<b>3.92 ± 1.11</b>
	Traditional (Random selection)	0.408 ± 0.122	0.32 ± 0.13	0.64 ± 0.22	3.42 ± 1.83



This result can be explained by a fact that the groups created by the reference method were homogenous and also the previous study [44] confirmed that students tend to create homogenous groups on the basis of their personality traits (e.g. an extrovert with another extrovert).

A one-way ANOVA was employed to determine an effect of the group formation method on the groups' overall evaluation. When collaborative skills were considered, we got p-value 0.0028,  $F(2, 251) = 6.03$ ,  $\eta^2 = 0.046$ . Similarly, for personal traits we got p-value 0.0267,  $F(2, 251) = 3.677$ ,  $\eta^2 = 0.028$ . Thus, the achieved results reveal that the application of the particular group formation method leads to statistically significant difference in the overall evaluation with the moderate effect size. A post-hoc Tukey-Kramer test showed that the differences between the mean of overall evaluation of groups created by the proposed method and other two means are significantly different, while the difference between means of randomly and reference groups is not significant.

Finally, we evaluated the iterative improvement of the proposed method during the experiment. In other words, we examined how well the proposed method was able to learn input matrices (the assignments of collaborative characteristics to students and mutual characteristics' compatibility) and thus improve group formation by utilizing provided feedback. A Pearson correlation between moving average of the overall evaluation (interval = 5 groups) and the order of iteration ( $r = 0.311$ ,  $p = 0.005$ ) pointed out the increasing performance of the proposed group formation method (all experiment settings and circumstances remained stable for the whole time of the experiment).

#### 6.4 Additional Analyses to Quantitative Experiment

The quantitative experiment provided us also with a possibility to gather valuable information about collaborative learning in the purely real-time learning environment which is still only very rare in the educational domain.

**Students.** First of all, we analyzed the degree of students' involvement in the experiment. The number of tasks solved during the experiment was generally diverse ( $\bar{x} = 5$ ,  $SD = 4.72$ ). Students with worse study results tend to solve more tasks than other students ( $r = 0.29$ ,  $p < 0.001$ ). On the other hand, their average evaluation is lower than average evaluation of the groups with better students ( $r = 0.25$ ,  $p = 0.001$ ). We can explain this result by the influence of motivation. Students, who actively participated in the experiment and achieved the most successful results, were rewarded. Despite this undesired negative influence, it is important to see that the better students achieved the better average evaluation.

**Groups.** The collaborative platform created 208 groups consisting of two members, and 46 groups consisting of three members during the experiment. Triads achieved a higher average overall evaluation ( $\bar{x} = 0.442$ ,  $SD = 0.109$ ) in comparison with couples ( $\bar{x} = 0.405$ ,  $SD = 0.125$ ), however, this difference is statistically insignificant, ANOVA:  $F(1, 252) = 3.411$ ,  $p = 0.0659$ ,  $\eta^2 = 0.0134$ . The higher evaluation rate was caused mainly by more intensive interaction which influences the high-level indicators of collaboration quality, such as argumentation and reaching consensus or

time and task management.

Additionally, we evaluated a correlation between teachers' manual evaluation and other automatically calculated indicators. The highest correlation was calculated for fluidity of collaboration ( $r = 0.35$ ,  $p < 0.001$ ), sustaining mutual understanding ( $r = 0.18$ ,  $p = 0.002$ ), argumentation and reaching consensus ( $r = 0.18$ ,  $p = 0.002$ ) and information exchanges for problem solving ( $r = 0.16$ ,  $p = 0.005$ ). We can derive several findings from these results. The more successful groups are those in which students are able to divide the assigned task into several partial problems and afterwards participate on their solving with approximately the same share. Furthermore, the quality of collaboration is also positively influenced by the content that is created to be clearly understandable by all group members. Especially, we positively judge the influence of argumentation and reaching consensus because students expressed their agreement and disagreement with the proposal of other group members. This fact is in the contrast with results of similar researches where students tend to avoid critical evaluation of other group members. However, creative conflict is considered as a very important aspect of successful collaborative learning [45].

**Activities.** Finally, we were interested in a correlation between students' activities (messages in the semi-structured discussion) and the quality of achieved results. We calculated the highest correlation between teachers' manual evaluation and following activities: write a praise ( $r = 0.28$ ,  $p < 0.001$ ), propose an action ( $r = 0.23$ ,  $p < 0.001$ ) and warn of a mistake ( $r = 0.20$ ,  $p < 0.001$ ). Based on the calculation of how many students use the defined messages, we can derive additional interesting findings. As the most students used accept the proposed action ( $n = 55$ ), write a praise ( $n = 47$ ) and propose an action ( $n = 40$ ), the collaboration process and the achieved results are positively influenced by students' self-regulation. Students are able to independently manage their collaboration, warn other members about eventual imperfections of the created solution and thus improve the result of their collaboration. We positively evaluate also the finding that students are able to motivate themselves mutually by writing a praise for a well created contribution to the overall solution.

## 7 DISCUSSION AND CONCLUSION

Nowadays, collaboration between learners is present in many web-based educational applications. This trend causes that we have to face new challenges. One of them is a study group composition, which plays an important role as it can significantly influence the process of collaborative learning. In spite of many existing methods to group formation, there are a lot of unresearched possibilities how to improve collaboration. We focused on one of them, namely on how to create dynamic short-term groups iteratively and automatically without student participation.

Our main contribution is the proposal of the novel method for automatic formation of dynamic groups based on Group Technology (GT) approach. In contrast to the existing methods for group formation based on GT approach, the proposed method is applied iteratively. This allows us



to take into consideration already achieved students' results and adjust the input parameters to provide better support during following collaboration. It means that we can start the group formation process with no or minimal information about learners and related characteristics. Our method automatically learns which collaborative characteristics are typical for students and which characteristics should be combined together to achieve more effective collaboration. It means that the proposed method can be characterized as a theory-free bottom-up approach.

We have successfully applied the proposed method in the collaborative platform PopCorm which provides students with the appropriate environment for effective communication and collaboration. It was also used as a tool to evaluate the proposed method during an experiment with 110 students. The results of the experiment show that the study groups created by the proposed method achieved the higher collaboration quality in comparison with the reference groups.

We identified many possibilities how to improve current design of the proposed method and its application in collaborative learning. We have not focused on task assignments to created groups in our work. It provides promising possibility how to further improve learners' collaboration because each group has different characteristics and different tasks are suitable in a particular moment of collaboration. Personalization of task assignment based on task's and group's attributes (e.g. knowledge of relevant domain terms which are necessary to achieve correct task solution) represents an interesting potential. Group recommendation principles can be employed here [46].

The design of our method is quite universal (especially due to the independence on particular characteristics) and thus the method has a potential to be applied in informal learning settings (e.g. in workspace) and also outside of the educational domain. In our further work, we plan to study how dynamic groups can support collaboration in knowledge sharing applications based on communities. More specifically, we are interested in Community Question Answering systems, such as Yahoo! Answers or Stack Overflow.

## ACKNOWLEDGMENT

This work was supported by grants No. VG1/0675/11, VG1/0971/11, KEGA 009STU-4/2014 and it is a partial result of the Research and Development Operational Program for the projects SMART, ITMS 26240120005 and SMART II, ITMS 26240120029, co-funded by ERDF.

The authors wish to thank colleagues from the Institute of Informatics and Software Engineering and all students (in particular members of PeWe group, [pewe.fiit.stuba.sk](http://pewe.fiit.stuba.sk)) for their invaluable discussions on the work presented in this paper. Special thanks deserve members of ALEF team for their direct contribution to tasks and educational materials used in our experiments. In addition, the authors would like to thank Milica Schraggeová, Danka Babincová and Katarína Babincová (Department of psychology, Comenius University in Bratislava) for preparing and evaluation NEO-FFI questionnaires for our bachelor students who participated in the experiments.

## REFERENCES

- [1] G. Stahl, T. Koschmann, and D. Suthers, "Computer-supported collaborative learning: An historical perspective," in *Cambridge handbook of the learning sciences*, 2006, pp. 409–426.
- [2] J. Vassileva, "Toward social learning environments," *IEEE Transactions on Learning Technologies*, vol. 1, no. 4, pp. 199–214, 2008.
- [3] C. Safran, D. Helic, and C. Gütl, "E-Learning practices and Web 2.0," in *Proc. of the Int. Conf. of Interactive Computer Aided Learning*, 2007, vol. 1, no. 1, pp. 1–9.
- [4] T. Daradoumis, M. Guitert, F. Gimenez, J. M. Marques, and T. Lloret, "Supporting the composition of effective virtual groups for collaborative learning," in *Proc. of Int. Conf. on Computers in Education*, 2002, vol. 1, pp. 332–336.
- [5] B. W. Tuckman, "Developmental sequence in small groups," *Psychological Bulletin*, vol. 63, no. 6, pp. 384–399, 1965.
- [6] B. Tuckman and M. Jensen, "Stages of small-group development revisited," *Group Organization Management*, vol. 2, no. 4, pp. 419–427, 1977.
- [7] D. L. Largent and C. Lüer, "'You Mean we Have to Work Together !?!': A study of the formation and interaction of programming teams in a college course setting," in *Proc. of the 6th int. workshop on Computing education research*, 2010, pp. 41–50.
- [8] F. Gimenez, M. Guitert, A. Daradoumis, J. M. Marques, and T. Lloret, "Cooperative Learning & Working in a Virtual Environment: a Case Study," in *Proc. of the 2nd European Distance Education Network, Research and policy in open and distance learning*, 2002, pp. 228–233.
- [9] T. R. Henry, "Creating effective student groups: An Introduction to groupformation.org," in *Proc. of the 44th ACM technical symposium on Computer science education - SIGCSE '13*, 2013, pp. 645–650.
- [10] K. Deibel, "Team formation methods for increasing interaction during in-class group work," *ACM SIGCSE Bulletin*, vol. 37, no. 3, p. 291, Sep. 2005.
- [11] A. Ounnas, H. Davis, and D. Millard, "A Framework for Semantic Group Formation," in *Proc. of 8th IEEE Int. Conf. on Advanced Learning Technologies*, 2008, pp. 34–38.
- [12] A. Gogoulou, E. Gouli, and G. Boas, "Forming homogeneous, heterogeneous and mixed groups of learners," in *Proc. of PING Workshop*, 2007, pp. 33–40.
- [13] D. Meyer, "OptAssign—A web-based tool for assigning students to groups," *Computers & Education*, vol. 53, no. 4, pp. 1104–1119, Dec. 2009.
- [14] M. Wessner and H.-R. Pfister, "Group formation in computer-supported collaborative learning," in *Proc. of the 2001 Int. ACM SIGGROUP Conf. on Supporting Group Work - GROUP '01*, 2001, pp. 24–31.
- [15] N. Rubens, M. Vilenius, and T. Okamoto, "Automatic Group Formation for Informal Collaborative Learning," in *Proc. of IEEE/WIC/ACM Int. Joint Conf. on Web Intelligence and Intelligent Agent Technology*, 2009, no. Figure 1, pp. 231–234.
- [16] I. Magnisalis, S. Demetriadis, and A. Karakostas, "Adaptive and Intelligent Systems for Collaborative Learning Support: A Review of the Field," *IEEE Transactions on Learning Technologies*, vol. 4, no. 1, pp. 5–20, Jan. 2011.
- [17] A. Dimitracopoulou, "Designing collaborative learning systems: current trends & future research agenda," in *Proc. of the conf. on Computer support for collaborative learning: the next 10 years!*, 2005, pp. 115–124.
- [18] M. Heinrich, F. J. Grüneberger, T. Springer, and M. Gaedke, "Reusable Awareness Widgets for Collaborative Web Applications—A Non-invasive Approach," in *Proc. of the 12th int. conf. on Web Engineering*, 2012, pp. 1–15.
- [19] C. Gutwin and S. Greenberg, "A descriptive framework of workspace awareness for real-time groupware," *Computer Supported Cooperative*

- Work (CSCW)*, vol. 11, no. 3, pp. 411–446, 2002.
- [20] N. M. Webb, "Task related verbal interaction and mathematics learning in small groups," *Journal for Research in Mathematics Education*, vol. 22, no. 5, pp. 366–389, 1991.
- [21] A. S. Wu, R. Farrell, and M. K. Singley, "Scaffolding group learning in a collaborative networked environment," in *CSCL '02 Proc. of the Conf. on Computer Support for Collaborative Learning: Foundations for a CSCL Community*, 2002, pp. 245–254.
- [22] J. Zumbach, J. Schönemann, and P. Reimann, "Analyzing and supporting collaboration in cooperative computer-mediated communication," in *Proc. of the conf. on Computer support for collaborative learning: the next 10 years!*, 2005, pp. 758–767.
- [23] P. Dillenbourg, "Over-scripting CSCL: The risks of blending collaborative learning with instructional design," in *Three worlds of CSCL. Can we support CSCL?*, P. A. Kirschner, Ed. Heerlen: Open Universiteit Nederland, 2002, pp. 61–91.
- [24] P. Dillenbourg and F. Hong, "The mechanics of CSCL macro scripts," *Int. Journal of Computer-Supported Collaborative Learning*, vol. 3, no. 1, pp. 5–23, Jan. 2008.
- [25] S. Demetriadis and A. Karakostas, "Introduction to Adaptive Collaboration Scripting Pedagogical and Technical Issues," in *Intelligent Collaborative e-Learning Systems and Applications*, Studies in, T. Daradoumis, S. Caballé, J. M. Marquès, and F. Xhafa, Eds. Berlin, Heidelberg: Springer, 2009, pp. 1–18.
- [26] A. Soller, A. Martínez, P. Jermann, and M. Muehlenbrock, "From Mirroring to Guiding: A Review of State of the Art Technology for Supporting Collaborative Learning," *Int. Journal of Artificial Intelligence in Education*, vol. 15, no. 4, pp. 261–290, Dec. 2005.
- [27] N. Rummel, H. Spada, and S. Hauser, "Learning to collaborate in a computer-mediated setting: observing a model beats learning from being scripted," in *Proc. of the 7th int. conf. on Learning sciences*, 2006, pp. 634–640.
- [28] I. Srba and M. Bieliková, "Encouragement of Collaborative Learning Based on Dynamic Groups," in *Proc. of the 7th European Conf. of Technology Enhanced Learning, EC-TEL 2012, Lecture Notes in Computer Science 7563*, 2012, vol. 7563/2012, pp. 432–437.
- [29] J. Moreno, D. a. Ovalle, and R. M. Vicari, "A genetic algorithm approach for group formation in collaborative learning considering multiple student characteristics," *Computers & Education*, vol. 58, no. 1, pp. 560–569, Jan. 2012.
- [30] Y. Lin, Y. Huang, and S. Cheng, "An automatic group composition system for composing collaborative learning groups using enhanced particle swarm optimization," *Computers & Education*, vol. 55, no. 4, pp. 1483–1493, 2010.
- [31] H. M. Selim, R. G. Askin, and A. J. Vakharia, "Cell formation in group technology: Review, evaluation and directions for future research," *Computers & Industrial Engineering*, vol. 34, no. 1, pp. 3–20, Jan. 1998.
- [32] M. Cocea and G. Magoulas, "Group formation for collaboration in exploratory learning using group technology techniques," in *Proc. of the 14th Int. Conf. on Knowledge-Based and Intelligent Information and Engineering Systems*, 2010, vol. 6277, pp. 103–113.
- [33] Y. a. Pollalis and G. Mavrommatis, "Using similarity measures for collaborating groups formation: A model for distance learning environments," *European Journal of Operational Research*, vol. 193, no. 2, pp. 626–636, Mar. 2009.
- [34] M. Cocea and G. D. Magoulas, "User behaviour-driven group formation through case-based reasoning and clustering," *Expert Systems with Applications*, vol. 39, no. 10, pp. 8756–8768, Aug. 2012.
- [35] L. E. Agustín-Blas, S. Salcedo-Sanz, E. G. Ortiz-García, A. Portilla-Figueras, Á. M. Pérez-Bellido, and S. Jiménez-Fernández, "Team formation based on group technology: A hybrid grouping genetic algorithm approach," *Computers & Operations Research*, vol. 38, no. 2, pp. 484–495, Feb. 2011.
- [36] A. Soller, "Supporting social interaction in an intelligent collaborative learning system," *Int. Journal of Artificial Intelligence in Education (IJAIED)*, vol. 12, pp. 40–62, 2001.
- [37] P. Jermann and D. Schneider, "Semi-Structured Interface in Collaborative Problem-Solving," in *Proc. of the First Swiss Workshop on Distributed and Parallel Systems*, 1997.
- [38] M. M. McManus and R. M. Aiken, "Monitoring computer-based collaborative problem solving," *Journal of Artificial Intelligence in Education*, vol. 6, no. 4, pp. 307–336, 1995.
- [39] J.-M. Burkhardt, F. Détienne, A.-M. Hébert, L. Perron, S. Safin, and P. Leclercq, "An approach to assess the quality of collaboration in technology-mediated design situations," in *European Conf. on Cognitive Ergonomics: Designing beyond the Product*, 2009.
- [40] H. Spada, A. Meier, N. Rummel, and S. Hauser, "A new method to assess the quality of collaborative process in CSCL," in *Proc. of the conf. on Computer support for collaborative learning: the next 10 years!*, 2005, pp. 622–631.
- [41] I. Srba and M. Bieliková, "Designing Learning Environments Based on Collaborative Content Creation," in *Proc. of Workshop on Collaborative Technologies for Working and Learning*, 2013, pp. 49–53.
- [42] M. Šimko, M. Barla, and M. Bieliková, "ALEF: A framework for adaptive web-based learning 2.0," *Key Competencies in the Knowledge Society*, vol. 324, pp. 367–378, 2010.
- [43] M. Bieliková, M. Šimko, M. Barla, J. Tvarožek, M. Labaj, R. Móro, I. Srba, and J. Ševcech, "ALEF: from Application to Platform for Adaptive Collaborative Learning," in *Recommender Systems for Technology Enhanced Learning: Research Trends & Applications*, Berlin, Heidelberg: Springer, 2014, pp. 195–225.
- [44] V. Hórreo and R. Carro, "Studying the impact of personality and group formation on learner performance," in *Groupware: Design, Implementation, and Use*, 2007, pp. 287–294.
- [45] T. Daradoumis, J. M. Marquès, M. Guitert, F. Giménez, and R. Segret, "Enabling novel methodologies to promote virtual collaborative study and learning in distance education," in *Proc. of the 20th World Conf. on Open Learning and Distance Education*, 2001, no. Timms.
- [46] M. Kompan and M. Bieliková, "Context-based Satisfaction Modelling for Personalized Recommendations," in *Proc. of the 8th Int. Workshop on Semantic and Social Media Adaptation and Personalization*, 2013, pp. 33–38.
- Ivan Srba** is currently a doctoral student at the Institute of Informatics and Software Engineering, Slovak University of Technology in Bratislava. He received master degree (with magna cum laude) in 2012 from the same university. His research interests are in the area of web-based systems which utilize concepts of collaboration and collective intelligence, in particularly he focuses on adaptive learning and knowledge sharing systems.
- Mária Bieliková** received her Master degree (with summa cum laude) in 1989 and her PhD. degree in 1995, both from the Slovak University of Technology in Bratislava. Since 2005, she has been a full professor, presently at the Institute of Informatics and Software Engineering at the same university. Her research interests are in the areas of web-based information systems, especially personalized context-aware web-based systems including user modeling, recommendation and collaboration support with special emphasis for the educational domain. She co-authored over 70 papers in international scientific journals. She served as an editor of more than 40 proceedings, eight of them published by Springer. She is a senior member of IEEE, a member of its Computer Society, senior member of ACM and member of International Society for Web Engineering.