The Use of Personality Traits to Enhance Theory-driven Group Formation

Rachel Carlos Duque Reis  
Universidade Federal de Viçosa  
Campus Rio Paranaíba - IEP  
rachel.reis@ufv.br

Kamila Takayama Lyra  
Universidade de São Paulo  
Campus São Carlos – ICMC/USP  
kalyra_03@usp.br

Clausius Duque Gonçalves Reis  
Universidade Federal de Viçosa  
Campus Rio Paranaíba - IEP  
clausius@ufv.br

Bruno Elias Penteado  
Universidade de São Paulo  
Campus São Carlos – ICMC/USP  
brunopenteado@usp.br

Seiji Isotani  
Universidade de São Paulo  
Campus São Carlos – ICMC/USP  
isotani@icmc.usp.br

Abstract

Group formation is an important and challenging element for designing successful CSCL scenarios. Despite efforts from the scientific community in developing more effective algorithms to support group formation processes, we still face problems related to learners’ resistance and demotivation towards group work. In this sense, diverse studies highlight the importance of considering learners’ personality traits to form groups, since this factor can influence students’ performance and induce diverse actions and behaviors in group work. Therefore, this paper presents G-FusionPT (Group Formation USing Ontology and Personality Trait), a group formation algorithm that support new learning roles, denominated Affective Collaborative Learning roles, based on relation between collaborative learning theories and students’ personality traits. The algorithm is based on a collaborative ontology to understand the learning theories (e.g., context, learning activities, group structure), and learners profile to understand learners’ needs (e.g., target/current knowledge/skill). To evaluate the algorithm, we used a 300 student simulated sample with varying group size (three, five, and seven members), and compared G-FusionPT results to other group formation algorithms: G-Fusion (based specifically on collaborative learning theories) and Random (no strategy or criterion). The results demonstrated the effectiveness of G-FusionPT against G-Fusion and Random algorithms, as it generated the highest average number of learners in well-formed groups and lowest average number of learners in unfit groups.

Keywords: CSCL, Group Formation, Algorithm, Personality Trait, Learning Theory.

1 Introduction

Computer Supported Collaborative Learning (CSCL) has been pointed out as an effective approach to maximize students’ learning gains. CSCL has been receiving more attention from scientific community and it has been used in different educational contexts, such as classrooms, e-learning/m-learning environments, learning management systems (LMS), and massive open online courses (MOOCs) (Brauer & Schmidt, 2012; Lin, Huang, & Cheng, 2010; Sinha, 2014; Srba & Bielikova, 2015; Amara et al., 2016).

In collaborative learning (CL), group formation poses as an important and complex element for designing successful CSCL scenarios (Inaba et al., 2000). According to Dillenbourg (2002), forming groups without any criteria or strategy (e.g., randomly) is usually ineffective, since there is no guarantee that such composition will trigger the expected interactions and learning among their members.

Literature reviews, developed by Cruz and Isotani (2014) and Borges et al. (2018), identified diverse solutions as computational algorithms, models, and tools to support the group formation process. These solutions mostly focus on matching different factors, such as learners’ characteristics, knowledge/skills, and context information to obtain the expected learning and behavior from a group. Isotani et al. (2009) present as a weakness of such solutions, the impossibility of justifying pedagogically the selection of participants to compose a group. To mitigate this problem, the authors developed a CL ontology, based on CL theories, that allows the formulation of different algorithms to support the creation of effective learning groups. The term effective, employed in Isotani et al. (2009), refers to the use of one or more CL theories to validate the success of group formation, with the use of essential factors to design CSCL scenarios such as CL roles. According to Inaba and Mizoguchi (2004), CL roles are a relevant resource to foment learners’ awareness during group work interactions and collaborations. Thus, to achieve the expected educational benefits, it is important to establish appropriate CL roles to each student, as well as to define their behavior in group work.

Although the important contributions of previous studies, researchers from CSCL still face the problem of students’ demotivation and resistance to group work (So & Brush, 2008; Caspi & Blau, 2011; Challco et al., 2015). So and Brush (2008) associate such resistance to students' sense of belonging and connectedness with other colleagues. According to the authors, this perception is influenced by emotional connections among students that may affect the quality of interactions and group results. In this sense, Jaques and Nunes (2012) discuss that considering only students’ skills in the group formation process does not ensure the participation of all its members. For instance, heterogeneous groups, in terms of social and affective characteristics, tend to spend more time in the resolution of socio-emotional conflicts than in solving the problem itself (Jaques & Nunes, 2012). Thus, CL researchers highlight the importance of considering learners’ personality traits to form groups, since such factor may influence students’ performance and induce diverse actions and behaviors in group work (Bradley & Herbert, 1997; Caropreso & Chen, 2003; Reis et al., 2016, 2018; Altanopoulou & Tselios, 2015).

Given the above, this work aims at extending the solution developed by Isotani et al. (2009) and presenting G-fusionPT (Group Formation USIng ONtology and Personality Traits), a group formation algorithm that support new CL roles, denominated Affective Collaborative Learning roles, that relate the collaborative learning theories to students’ personality traits (Reis et al., 2016). To assess the algorithm’s viability, this paper also presents a visual tool, namely VISO-EGO (VISual t0ol for Effective Group fOrmation), to run the G-FusionPT algorithm. In addition, we developed a simulation with 300 students comparing the G-FusionPT results with two other group formation algorithms, namely G-Fusion (based specifically on collaborative learning theories) and Random (no strategy or criterion). As a main result, G-FusionPT proved to be more effective than the G-Fusion and Random algorithms.

This paper is organized in eight sections: In Section 2, we offer an explanation about the main concepts used in this work. In Section 3, we describe a group formation algorithm, named G-FusionPT. In Section 4, we present a computational system to execute G-FusionPT algorithm. In Section 5, we show the simulation results based on real-world data to evaluate G-FusionPT, and compare it with two other group formation algorithms, namely G-Fusion and Random. In Section 6, we discuss the results, and in Section 7, we present the literature review about algorithms to support the group formation process involving students’ personality trait. Finally, in Section 8, we show the conclusions and future works.

2 Fundamental Concepts

This section introduces the main concepts related to the development of this work.
2.1 Learning Group Formation

Learning group formation represents the basic structure for the creation of educational scenarios that promote collaboration between two or more people (Isotani et al., 2009). In general, strategies to group formation are based on three main configurations: (1) at random, without any selection criteria; (2) with self-selection, where the students select who they would like to work with, or (3) by teacher selection, using a specific criteria (Barkley, Cross, & Major, 2005).

CSCL researches highlight their concern regarding random selection and self-selection approaches, since they can result in unequal learner participation. For instance, group members working at a different pace, off-task behavior, among others (Barkley, Cross, & Major, 2005; Inaba & Mizoguchi, 2008). In order to minimize such scenarios, many scientific studies have investigated different criteria (factors and strategies) to form groups more capable to achieve desired learning goals. A literature review about group formation in CSCL, developed by Borges et al. (2008), found in 106 papers three main factors: “knowledge level” (59 studies), “skill-based group” (33 studies) and “learners’ role” (31 studies). Other factors found in less than five studies were “social relationship”, “behavioral traits”, “personality”, and “cultural aspects”, among others.

2.2 Personality Trait

In the literature there are many different definitions for the term personality (Nunes, 2008). One of them is related to “trait approach”, allowing individuals to be psychologically differentiated by the use of measurable traits, called personality traits (Allport, 1927). Allport, in his studies, considered personality as a continuous process of growth and development. So, based on Allport viewpoint, in this work personality trait is defined as behavioral tendencies that cause individuals to behave in specific ways based on the context. The personality trait intensity varies according to each person and generally is related to the level (high, medium or low) in which one characteristic is expressed by an individual.

Table 1 presents some characteristics expressed in individuals with high and low levels in psychoticism personality trait, referred to this work as Mental Rigidity. For a better understanding, imagine two people with different levels of mental rigidity, one with high level and other with low level. The person with high level tends to express “aggressiveness”, “lack of empathy” and “hostility” with more intensity than one with low level.

2.3 Affective Collaborative Learning Role

An Affective Collaborative Learning role (ACL role) defines a new kind of CL role that matches collaborative learning theories to students’ personality traits. It is worth mentioning that
the process for creating new ACL roles allows the combination of any personality traits with any CL theories (Reis et al., 2016).

Table 2 presents some examples of ACL roles based on the combination between Mental Rigidity personality trait (Eysenck & Eysenck, 1975) and Peer Tutor/Peer Tutee CL roles, based on Peer Tutoring CL theory (Endlsey, 1980).

Table 2: Examples of ACL roles and their characteristics. Source: Reis (2019).

<table>
<thead>
<tr>
<th>ACL Roles</th>
<th>Prerequisites</th>
<th>Personality trait characteristic that may be threats</th>
<th>Behavior (problem)</th>
<th>Expected Benefits Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rigid peer tutor</td>
<td>* Having the target knowledge (1); * Having a <em>high level</em> of mental rigidity (2); - Misunderstanding the knowledge (3); - Not having experience in using the target knowledge.</td>
<td>non-empathy and aggressiveness</td>
<td>Tutoring (difficulty to transmit knowledge)</td>
<td>Improve the knowledge about the content. (acquisition of tuning)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non_rigid peer tutor</td>
<td>* Having the target knowledge; * Having a <em>low level</em> of mental rigidity; - Misunderstanding the knowledge; - Not having experience in using the target knowledge.</td>
<td>-</td>
<td>Tutoring</td>
<td></td>
</tr>
<tr>
<td>Rigid peer tutee</td>
<td>* Having a <em>high level</em> of mental rigidity; - Not having the target knowledge.</td>
<td>non-empathy and aggressiveness</td>
<td>Passive learning. (difficulty for following colleague's guidelines)</td>
<td>Acquisition of new knowledge. (acquisition of accretion)</td>
</tr>
<tr>
<td>Non_rigid peer tutee</td>
<td>* Having a <em>low level</em> of mental rigidity; - Not having the target knowledge.</td>
<td>-</td>
<td>Passive learning</td>
<td></td>
</tr>
</tbody>
</table>

Each ACL role has a prescribed behavior defined by the CL theory. For example, a student who performs the Rigid peer tutor role (column 1, Table 2) must adopt the behavior of Tutoring (column 4) with the purpose of explaining something he/she already knows to other learners. However, the characteristics of non-empathy and aggressiveness (column 3) may cause the behavior problem of difficulty to transmit the knowledge (column 4). Beside the relationship between ACL roles and the necessary behaviors to achieve educational benefits, it is important to specify the prerequisites (column 2), i.e., necessary ‘*’ and desired ‘-’ conditions to play an ACL role. The necessary conditions are essential for a student to play an ACL role. If the student does not attend these conditions, he/she cannot play the ACL role appropriately. The desired conditions define prerequisites that students must satisfy to ensure the educational benefits (column 5). For example, a student can only play the Rigid peer tutor role (necessary condition) if he/she: (1) has the knowledge about the content that will be covered, and (2) has a high level of mental rigidity personality trait. On the other hand, to ensure educational benefits, in this example, it is desirable (desired conditions) that the student who will play the Rigid peer tutee role: (3) misunderstand some parts of content that will be covered, and (4) has no experience in using the knowledge.

The Non_rigid peer tutor role differs from Rigid peer tutor role only in the personality trait prerequisites (column 2), which states that the learner must have “a low level of mental rigidity personality trait” (necessary condition). In addition, this ACL role has none personality trait characteristic that may cause problems in students’ interaction (column 3).
Other examples of ACL roles can be found in the study developed by Reis et al. (2016, 2017). In Reis et al. (2016) the ACL roles were based on a combination between introvert/extrovert personality traits and Anchored instructor/Problem holder CL roles, based on Anchored Instructor CL theory (CTGV, 1992). For example, Introverted anchored instructor, Extroverted anchored instructor, Introverted problem holder and Extroverted problem holder.

2.4 Ontology

According to Borst (1997, p. 12) an ontology is a “formal and explicit specification of a shared conceptualization”. The formal term refers to something that is understandable by computers; explicit specification represents the concepts, properties, relations, functions, restrictions and axioms of a domain; and shared conceptualization means consensual knowledge among people and/or computational applications.

In the CL context, Inaba et al. (2000) was one of the first studies to use ontologies for describing CL scenarios. Based on that, other studies were conducted for the creation of models and strategies to support, for example, the planning of collaborative activities, the formation of learning groups, and the engagement/interaction among students (Inaba & Mizoguchi, 2004; Isotani et al., 2009; Isotani et al., 2013; Challco et al., 2015; Reis et al., 2016).

The use of ontologies for modeling CL scenarios has a lot of benefits. For example, (1) representing a knowledge domain with potential of reuse, (2) formalization of a domain, eliminating contradictions, ambiguities and inconsistencies; (3) knowledge sharing between people and/or computational applications (Borst, 1997; Mizoguchi, 2003).

In this work, we use an up-to-date version of CL Ontology (Reis, 2019, p. 214), originally proposed by Isotani et al. (2009). More details about the CL Ontology are shown in Section 3.1.2.

3 Group Formation Algorithm: G-FusionPT

This session presents G-FusionPT, an open source algorithm, implemented in PHP language that extends the solution developed by Isotani et al. (2009), involving students’ personality traits. In Subsection 3.1, we present the main concepts used in G-FusionPT. Next, in Subsection 3.2, we describe the algorithm step by step, and finally, in Subsection 3.3, we present a pseudocode of the G-FusionPT algorithm.

3.1 Main Concepts

The main concepts used in G-FusionPT algorithm were extracted from two sources: learners’ profile and an up-to-date version of CL Ontology (Reis, 2019), originally proposed by Isotani et al. (2009).

3.1.1 Learners’ Profile

Learner’s profile is defined by expression $V_0 \rightarrow V_1$, where $V_0$ refers to the student’s learning initial state, denominated in this paper as current learning. $V_1$ represents the student’s learning final state, named target learning (Figure 1).
Students’ current learning \((V_0)\) is composed of three parameters as follows: \(V_0 = (x_0, y_0, z)\). The first parameter \((x_0)\) represents the current state of learning and it is characterized by four stages \((x_0 = \{0...3\})\) concerning the process of knowledge acquisition (Rumelhart & Norman, 1978): (0) Nothing, (1) Accretion, (2) Tuning, and (3) Restructuring. “Nothing” means the student has no knowledge about some specific information. “Accretion” refers to the learning stage where the student acquires and accumulates new information, but he/she is not able to use it yet. “Tuning” represents a learning stage where the student interprets and applies the information in a specific context. “Restructuring” is the stage where the learner generates new information based on the relation between knowledge acquired and his/her capacity of interpretation.

The second parameter \((y_0)\) represents the current state of skill and it is composed by five stages \((y_0 = \{0...4\})\) concerning the skill development (Anderson, 1982): (0) Nothing, (1) Rough-cognitive stage, (2) Explanatory-cognitive stage, (3) Associative stage, and (4) Autonomous stage. “Nothing” means the student never was introduced to the skill. In the “cognitive” stages (rough and explanatory), the individual presents a desired behavior or, at least, some rough approximation thereof. In the “associative stage” the individual improves the desired behavior through practice, identifying and eliminating some mistakes presented initially. The “autonomous stage” is the progressive improvement of the desired behavior, in which the individual is able to perform the desired behavior with accuracy and quickly.

Finally, the third parameter \((z)\) represents student’ personality traits, which can be any personality trait and its different levels (e.g., low, medium, high). For example, extroversion, introversion, neuroticism, mental rigidity, among others.

Students’ target learning \((V_1)\) has two parameters as follows: \(V_1 = (x_1, y_1)\). \(X_1\) and \(y_1\) represent, respectively, the same stages of knowledge acquisition and skill development previously presented. Although, \(V_1\) represents the expected level of knowledge to be achieved by students at the end of a collaborative session.

### 3.1.2 CL Ontology

Regarding the CL Ontology, there are three main concepts to consider in G-FusionPT: “Collaborative Learning Theory (CL Theory)”, illustrated by Figure 2(a); “Learning Strategy”, illustrated by Figure 2(b); and “Affective Collaborative Learning role (ACL role)”, illustrated by Figure 2(c).
Figure 2: Part of CL Ontology with the main concepts for group formation.

The “CL theory” (Figure 2(a)) is a conceptual scheme that explains the learning process in a collaborative scenario and can be used to provide some essential conditions to make learning more effective (Isotani, 2009). Some examples of CL theories are Anchored Instruction (CTGV, 1992), Distributed Cognition (Salomon, 1993), Peer Tutoring (Endlsey, 1980), among others. CL theories enable to design pedagogical CL scenarios that specify, for example, the context, activities, learning strategies, learners’ roles, target goals (knowledge, skill), and other elements that influence the learning process.

The “learning strategy” (Figure 2(b)) defines how a specific learner (Learner1) may interact with another learner (Learner2) in order to achieve the strategy benefits. Such benefits are based on knowledge acquisition (accretion, tuning, and restructuring) and skill development (rough-cognitive stage, explanatory-cognitive stage, associative stage, and autonomous stage). One example of a learning strategy for scenarios based on Peer Tutoring CL theory and Mental Rigidity personality trait is learning by explanation combined with dynamics for promoting students’ empathy (Reis, 2019, p. 213).

Finally, the “ACL roles” (Figure 2(c)) are essential in the process to form appropriate groups of learners, since they represent the role to be performed for each group member during a collaborative session. As explained in Section 2.3, ACL roles are characterized by matching CL roles (e.g., Peer Tutor/Peer Tutee, Anchored Instructor/Problem Holder) to learners’ personality traits (e.g., mental rigidity, extroversion/introversion). An ACL is composed of necessary and desired conditions that specify who may or may not play the role. As shown by Figure 2(c), such conditions are based on learner’s personality trait, his/her current knowledge (nothing, accretion, tuning, and restructuring), and his/her current skill (nothing, rough-cognitive stage, explanatory-cognitive stage, associative stage, and autonomous stage). Examples of ACL roles are Rigid/Non_rigid peer tutor and Rigid/Non_rigid peer tutee (Reis, 2019, p. 213), Extroverted/Introverted anchored instructor and Extroverted/Introverted problem holder (Reis et al., 2015b, 2016), Stable/Unstable extroverted (Reis et al., 2017).

3.2 Steps

We have many possibilities to form groups combining the CL ontology (Figure 2) and learners’ profile (Figure 1). In this paper we are going to explore one possibility that considers learners’ target goals (knowledge and skill). In this section, we summarize the algorithm steps considering two phases: candidates’ selection (Figure 3) and group formation (Figure 4).
1. Candidates’ selection (Figure 3): the algorithm takes as input a list of Learners selected by a teacher, and a list of CL Theories from CL Ontology (Figure 3(a)). Next, the algorithm verifies whether the target learning from the learner’s profile matches to strategy benefits in CL Ontology (Figure 3(b)). If so, the algorithm verifies whether the current learning from the learner’s profile matches the ACL role’s necessary conditions in CL Ontology (Figure 3(c)). If so, the algorithm generates a list of member candidates for each learning theory with students.

2. Group formation (Figure 4): the algorithm takes as input a list of member candidates generated previously (phase 1) and a group size, provided by the teacher (Figure 4(a)). Based on candidates’ list for a given CL theory, and considering that A1 is the student in analysis, the algorithm selects an ACL role, named [L], for the learner A1 (Figure 4(b)) and an ACL role, named [R], that can interact with [L] (Figure 4(c)). If [L] and [R] correspond to the same ACL roles (Figure 4(d)), the student A1 will be added to the group (Figure 4(e)). If [L] and [R] are different ACL roles, the algorithm checks whether there is a space for [L] in the group (Figure 4(f)). If so, student A1 will be added to the group (Figure 4(e)). Otherwise, student A1 will be added to the list of orphan students (Figure 4(g)), which consists of learners who have not been added to any learning group.
3.3 Pseudocode

Table 3 and 4 present G-FusionPT pseudo-algorithms to, respectively, generate the list of member candidates and, based on such list to form the learning groups. This algorithm is an adaptation of the one proposed by Isotani et al. (2009).

Table 3: Pseudocode to generate the list of candidates.

```plaintext
Generate_CandidatesList (learner U, CLOntology CLO)
  // For each learner U from the list of learners
  For each learner U do
    // For each theory from the CL ontology
    For each theory T in CLO do
      // For each strategy from the CL theory
      For each strategy ST in T do
        U.I-goal // Obtain learner “target learning” from U profile.
        ST.I-goal // Obtain “strategy benefits” when using the strategy
        // Verifies if “target learning” matches to “strategy benefits”
        If U.I-goal = ST.I-goal then
          U.conditions // Obtain student “current learning” from L profile
          ACL.requirements // Obtain “necessary condition” to play an (ACL) role
          // Verifies if students’ current learning fulfill the necessary
          // conditions to play ACL role (R)
          If canPlayACLRole(U.conditions, ACL.requirements) then
            addCandidateList(U, T, ST, ACL) // Add U learner to candidate list
```

Figure 4: Steps of G-FusionPT algorithm to form the learning groups (Phase 2).
Table 4: Pseudocode to form the learning groups.

```plaintext
Group_Formation (CandidateList List, GroupSize S)
// For each CL theory in candidates list
For each theory T in the List do
  // For each learning strategy of CL theory.
  For each strategy ST in T do
    // For each learner using strategy ST.
    For each learner U in ST do
      I-role = U.ACL // Obtain the ACL role for learner U.
      You-role = getRole(U.ACL, ST) // Obtain other ACL role that may interact
      // with I-role according to strategy ST
      // Group members will play the same role
      If I-role = You-role then
        addLearner(U, I-role, ST, i) // Add learner L to group i playing I-role
        // using the strategy ST
      // Group members will play different roles
      Else If I-role != You-role then
        // If I-role fits in group i.
        If hasSpaceFor(I-role, i) // Add learner L to group i playing
        addLearner(U, I-role, ST, i) // I-role using the strategy ST.
      // Leftover learners that were not included in any group i after check
      // all CL theories in the List
      If hasOrphanLearners then
        // Run others group formation algorithms. In our study, we ran
        // G-Fusion and Random algorithms (presented in Section 5).

4 VISÓ-EGO Tool

VISÓ-EGO (VISual tOol for Effective Group fOrmation) is a computational system designed to run G-FusionPT and compare its results with Random and G-Fusion algorithms. In Figure 5 the menu is presented with options to execute each algorithm individually, or run a comparative mode in “Compare Algorithms”, where all three algorithms are executed simultaneously.

Figure 5: Group formation menu in VISÓ-EGO tool.

The option for comparing the three algorithms performs an analysis of the results based on two elements: distribution of CL roles and combination of personality traits. For example, CL roles in Table 5 list the requirements to form groups based on Peer Tutoring CL theory (Endsley, 1980).

Table 5: Requirements to form groups based on Peer Tutoring CL theory (Endsley, 1980).

<table>
<thead>
<tr>
<th>CL role</th>
<th>Members</th>
<th>Members</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Endsley, 1980)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Regarding personality trait, Table 6 shows the recommendation (yes or maybe) for the interaction between students with Mental Rigidity personality trait, when they play the CL roles of Peer Tutor and Peer Tutee (Endsley, 1980). Such recommendation was established based on experimental results obtained by Reis (2019, p. 63).

Table 6: Recommendation (yes or maybe) for the interaction between students with Mental Rigidity personality trait.

<table>
<thead>
<tr>
<th>Student 1 – Peer Tutor</th>
<th>Student 2 – Peer Tutee</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low level</td>
<td>Low Level</td>
<td>Yes</td>
</tr>
<tr>
<td>Low level</td>
<td>Medium level</td>
<td>Maybe</td>
</tr>
<tr>
<td>Medium level</td>
<td>Low level</td>
<td>Maybe</td>
</tr>
<tr>
<td>Medium level</td>
<td>Medium level</td>
<td>Maybe</td>
</tr>
</tbody>
</table>

When selecting the option for comparing the three algorithms, the processing results of the VISO-EGO tool produce a visualization based on circles, where each circle represents a formed group with smaller circles inside representing its members. Figure 6 shows an example of visualization output for a group size of three students. Group validity, at the bottom, is represented by the color range from red or thick black border (0% valid – unfit group) to green (100% valid – well-formed group). Below each algorithm results, the number of formed groups is presented, divided into validity levels.

Figure 6: Comparative results among Random, G-Fusion, and G-FusionPT algorithms, with groups formed by three members.

Whenever a user clicks in any of the groups formed in Figure 6, the selected group is expanded, providing information about the CL role and personality trait of its members (Figure 7). The color of each member, represented as smaller circles, also varies according to CL roles and/or personality traits validity. For example:

- The white color indicates a correct combination of CL roles and personality trait, as shown in every group members of G-FusionPT algorithm in Figure 7;
• The pink color represents a problem either in CL roles distribution or in personality trait combination. This can be observed in two members of Random algorithm, in Figure 7, with CL role problem, i.e., both are playing the Peer Tutor role;
• The gray color indicates the combination could be adequate. For instance, in G-Fusion algorithm (Figure 7), the Tutor member with medium level (RM) of mental rigidity has some chance (~50%) to work well with low level Tutees (RM--; or RM-).

Figure 7: Examples of group information in VISO-EGO tool.

5 G-FusionPT Evaluation

This section presents a simulation based on real-world data to measure G-FusionPT effectiveness. This simulation compares G-FusionPT with two other group formation algorithms, namely “Random” and “G-Fusion”. Random algorithm does not use strategies or criterions to create the learning groups, and G-Fusion algorithm (Isotani et al., 2009) forms the learning groups based specifically on CL theories.

5.1 Simulation Design

In order to obtain a larger sample of students, we performed a Monte Carlo simulation modeling the variables of interest: acquisition of knowledge and mental rigidity level. This simulation is a scientific tool used for experimentations which are too time-consuming, costly, or impractical to obtain real data (Harrison, 2010). In a Monte Carlo simulation, values are sampled at random according to the input probability distributions. Thus, data provides a much more comprehensive view of what may happen and how likely it is to happen. The probability distribution of these variables (acquisition of knowledge and mental rigidity level), shown in Table 7, was inferred from real-world data, i.e., from experimental study developed by Reis (2019, p. 63). Thus, they are a representative sample of the population under study. For each of the two variables, we simulated possible values for 300 students. The sample size of 300 students was determined using the Cochran’s formula (Bartlett et al., 2001), with a confidence level of 95%.

| Table 7: Probability distribution of variables (knowledge and personality trait). |
|------------------|------------------|------------------|
| Acquisition of Knowledge | Mental rigidity level |
| Nothing         | 36.60%           | Low              | 48.90%           |
| Accretion       | 62.80%           | Medium           | 40.50%           |
We used these 300 students’ simulated samples as input to compare G-FusionPT to Random and G-Fusion algorithms. We considered three different scenarios that vary on the number of students in the groups. Thus, group sizes vary from three, five, and seven students. Although the algorithms allow the creation of groups of any size, bigger or equal 2 (two), we have chosen these three scenarios to verify how the algorithms behave with an increasing of group members number. We run each algorithm 30 times for each group size in order to obtain statically analyzable data. Considering each scenario individually, we calculated the algorithms effectiveness of each group formation algorithm based on the following measures:

- the average number of learners in well-formed groups (100% valid),
- the average number of learners in unfit groups (0% valid),
- the average number of orphan learners. Orphan learners represent those individuals who are not assigned to any group.

5.2 Results and Interpretation

Figures 8, 9, and 10 shows the algorithms results (Random, G-Fusion, G-FusionPT), using VISO-EGO tool, considering the three scenarios previously presented, varying group size to three (gs3), five (gs5), and seven (gs7) members. Based on Figures 8, 9 and 10, we can observe that in all cases, the average number of learners in “unfit groups” is always higher than learners in “well-formed groups”. This result was expected, since only 48.90% students had a low level of mental rigidity personality trait (Table 7), and just groups composed by students in this level were considered well-formed (Table 6).

As shown in Figures 8, 9 and 10, the G-FusionPT algorithm had the lowest average number of learners in “unfit groups” ($\overline{gs3} = 126$, $\overline{gs5} = 155$, $\overline{gs7} = 163.33$), while the Random algorithm had the highest average ($\overline{gs3} = 218.60$, $\overline{gs5} = 273.83$, $\overline{gs7} = 288.17$). This difference has statistical significance\(^1\) in all three scenarios: groups of size 3 ($p < 0.000$), 5 ($p < 0.000$), and 7 ($p < 0.000$), where $p$ is the significance value.

The resulting values for the Random and G-FusionPT algorithms, regarding learners in “well-formed groups”, present opposite numbers in Figures 8, 9 and 10. The Random algorithm had the lowest average number of students in “well-formed group” ($\overline{gs3} = 9.60$, $\overline{gs5} = 1.50$, $\overline{gs7} = 0.00$), while G-FusionPT had the highest average ($\overline{gs3} = 93.00$, $\overline{gs5} = 75.00$, $\overline{gs7} = 70.00$). This difference is also statistically significant in all three scenarios: groups of size 3 ($p < 0.000$), 5 ($p = 0.011$), and 7 ($p < 0.000$).

Regarding the G-Fusion algorithm, the average number of learners in “well-formed groups” ($gs3 = 24.40$, $gs5 = 3.67$, $gs7 = 0.93$) is statistically lower than G-FusionPT for groups of size 3 ($p < 0.000$), 5 ($p < 0.000$), and 7 ($p < 0.000$). On the other hand, the average number of learners in “unfit group”, in both algorithms, was similar and did not have statistical difference in none of three scenarios: groups of size 3 ($p = 0.186$), 5 ($p = 0.842$), and 7 ($p = 0.186$).

\(^{1}\) Confidence level at 95%, using the statistical test ANOVA (Analysis of Variance).
Another aspect observed in Figures 8, 9 and 10 is the decrease of learners in “well-formed groups”, as well as an increase of learners in “unfit groups” with the growth of member numbers in each group. This can be justified by the difficulty of matching learners’ roles, as we included personality traits as a new prerequisite that students must satisfy in order to play an ACL role. Thus, this new factor increased the difficulty of producing valid groups with an increase in the probability of creating unfit groups.

Concerning the orphan learners, all three algorithms had an average number lower than seven learners. As the Random algorithm has no strategy or criterion to form groups, we calculate the number of orphan learners as following:

\[ \text{Orphan learner} = \frac{300}{\text{group size}} \]

Thus, all students were assigned to a group when the group size was 3 (Figure 8) and 5 (Figure 9). On the other hand, when the group size was 7 (Figure 10), six learners were left out.

Both, the G-Fusion and G-FusionPT algorithms, had the same average number of orphan learners in all three scenarios (Figures 8, 9 and 10). Additionally, such value was bigger than in the Random algorithm, because the number of students left out was less than the chosen group size. Thus, the algorithms G-Fusion and G-FusionPT were not able to form another group. For example, before forming the groups shown in Figure 6, the number of learners in the list of candidates to Peer Tutoring CL theory was 299, as one student did not meet the prerequisites to play its role (peer tutor or peer tutee). So, the G-Fusion and G-FusionPT algorithms formed 99 groups (Figure 6 - #Groups), instead of 100, with 2 (two) students being left out. In this case, the recommendation is to run the algorithms individually (Figure 5) for such students providing a smaller group size number. Another option is to apply individual learning for each of them.
6 Discussion of Results

In order to evaluate the algorithm’s effectiveness, the results presented in Section 5 just consider the average number of learners in well-formed groups (100% valid) and the average number of learners in unfit groups (0% valid). Regarding the learners in unfit groups (0% valid), some reasons may justify such formation. For example, (1) more than one student playing Peer Tutor role in a group, (2) members with high levels of Mental Rigidity personality trait. In the second case, some studies in the literature have observed, based on experimental studies, that students with high levels of Mental Rigidity generally present low academic performance in individual
learning (Heaven et al., 2007; Petrides et al., 2005). In collaborative learning, Reis (2019, p. 80) verified that groups based on Peer Tutoring CL theory (Endsley, 1980) and formed by members with high levels of Mental Rigidity resulted in less learning. However, the author discusses the importance of the creation of collaborative scenarios using other CL theories, for example, Distributed Cognition (Salomon, 1993), since it explores the creativity of students. As shown in Table 1, “being creative” is a characteristic of learners with high levels of Mental Rigidity.

But, what happens to other groups with validity between 0% and 100%? In Figure 6, such groups are represented by colors: “dark orange” (~16% valid), “light orange” (~33% valid), “yellow” (~50 valid) and “light green” (~66% valid). In order to ensure the educational benefits of group members during a collaborative session, it is important to define strategies to avoid behavior problems caused by some characteristics of individuals with high levels of Mental Rigidity personality trait. According to Cravens-Brown (2002), this trait is a predictor of problems related to an aggressive conduct, although the intensity and effect depend on the context where the individual is immersed. On the other hand, Rodrigues and Silva (2012) discuss the importance of empathy to inhibition of aggression during students’ interaction. Thus, to deal with behavior problems caused by characteristics such as the lack of empathy and aggressiveness, in students who perform the Rigid peer tutor or Rigid peer tutee CL roles (Table 2, column 3), CL researchers (Reis, 2019, p. 213) suggest a strategy based on dynamics for promoting students’ empathy.

Another example of strategy is the “socio-cognitive conflict technique” (Buchs & Butera, 2013). Such strategy consists of creating “disturbing” situations where the student self identifies contradictions in their statements, reformulating the reasoning, without the correct answer being presented. Pacheco and Sisto (2003) applied this strategy in group work, where they observed that learners with high levels of Mental Rigidity personality trait had better learning performance. The authors believe that the difficulty of socialization, lack of empathy and insensitivity, each being characteristics of individuals with a high level of Mental Rigidity, helped such learners to reflect on the conflict and overcome their contradictions.

Therefore, if some of these learning strategies (e.g., dynamics for promoting empathy, socio-cognitive conflict technique) are used in a collaborative session supported by a CL theory, the learners in groups with validity lower than 100%, have a better chance of achieving their learning goals. It is important to highlight that such strategies are applied after the group formation step. Thus, the G-FusionPT algorithm does not consider such strategies to form the learning groups.

7 Related Works

In literature, different expressions are found referring to “group/team formation”. For instance, group/team creation, group/team design, group/team composition and group/team organization. Based on such terms, Cruz and Isotani (2014) performed a systematic mapping of literature, where they claim 91.6% of studies either propose or implement an algorithm to support group formation processes. From those, only 4.2% are based on personality traits diversity to compose learning groups (Graf & Bekele, 2006; Gogoulou et al., 2007).

Graf and Bekele (2006) developed an “Ant Colony Optimization” algorithm to maximize the heterogeneity of groups, where they created a mathematical model based on two characteristics: students’ personality traits and their learning performance. The authors’ assumption is that heterogeneous groups increase students’ learning gains. By using the same characteristics, Gogoulou et al. (2007) proposed a tool that implements three different group formation algorithms: (1) purely homogeneous in both characteristics, (2) purely heterogeneous in both characteristics, and (3) using genetic algorithm to create homogeneous and
heterogeneous, as well as mixed groups (homogenous in personality traits and heterogeneous in learning performance). The genetic algorithm approach presented better results in forming homogeneous and mixed groups.

Other studies collected by Cruz and Isotani (2014) presented flexible group formation algorithms, which allows different types and numbers of attributes as input (Craig, Horton, & Pitt, 2010; Abnar, Orooji, & Taghiyareh, 2012; Cadavid, Ovalle, & Vicari, 2012). In particular, the experimental study developed by Abnar, Orooji and Taghiyareh (2012) considered as input the learners' personality traits and learning style. The results showed that the quality of groups formed by the algorithm was better than groups created using the random algorithm.

Reis et al. (2015, 2015a, 2018) and Amara et al. (2016) also developed literature reviews for the CSCL context. Reis et al. (2015) developed a systematic mapping to investigate the use of affective states, such as emotion and mood, and personality trait in CSCL environment. From 31 selected studies, only two (Grand & Bekele, 2006; Nunes, 2010) consider students’ personality traits in the algorithms for supporting group formation. Nunes (2010) proposes a recommender system, named “Group Recommender”, that aims at creating learning groups based exclusively on students’ personality traits (neuroticism, extraversion, openness, agreeableness, and conscientiousness). The Group Recommender system used a KNN (K-Nearest Neighbors) data mining algorithm to suggest learners with similar personality traits.

Reis et al. (2018) extended the previous review and found three other studies (Shin-ike, and Lima, 2008, 2009; Farhangian et al. 2015) to support group formation based on prediction models. Shin-ike and Lima (2008, 2009) observed the use of neural networks and genetic algorithms for the prediction of learning results in group work, based on their personality trait. The solution receives as input information about students’ academic background, their personality traits and correctness rate of tests performed individually. Farhangian et al. (2015) designed a simulation model based on agents for the prediction of students' performance. The results showed that the approach can be a low-cost tool for teachers and researchers to understand the impact of different factors (e.g., personality, knowledge and skill) on group formation.

The systematic literature review carried out by Amara et al. (2016) however, examined solutions for group formation in the mobile CSCL context. From 12 collected studies, four proposed algorithms to support group formation, but none took the students’ personality traits into account.

Other studies using personality traits to compose learning groups are: Yan and Chen (2013), Garcia-Vélez et al. (2016), and Konert et al. (2016). Yan and Chen (2013) used a clustering algorithm, denominated “Chameleon algorithm”, to create groups of up to eight students with maximal similarity of learning interest. Such similarity is based on six factors that compose the personality, these being the learning interest, learning capability, learning style, learning activity, sex and age. Garcia-Vélez et al. (2016) proposed a solution that uses the concepts of data mining in students’ personality traits and an algorithm, named “Cortical Learning Algorithm”, to form sporadic groups. This algorithm uses concepts of machine learning (e.g., Bayesian network, neural network) to find the better composition of students. Finally, Konert et al. (2016) presented MoodlePeers, a plugin for Moodle that implements “GroupAL” algorithm to form learning groups. The plugin generates dynamic questionnaires (personality traits, team orientation, motivation, prior knowledge) according to the desired learning scenario. Results showed that groups formed by this plugin had better performance and commitment than groups formed randomly.

Most group formation algorithms, presented in this section, focus on improving group diversity (e.g., homogeneous, heterogeneous, mixed) in terms of personality traits, using
different group formation strategies (e.g., mathematical model, genetic algorithm, data mining algorithm). In addition, other studies have proposed prediction models or flexible algorithms in terms of type and number of attributes for creating the learning groups. A summary of related works is shown in Table 6.

Although the valuable contributions of the previously presented studies, we identify some limitations. For example, (1) the lack of studies using students’ personality traits with pedagogical basis to justify the composition of groups in CSCL environments, (2) lack of studies that propose algorithms to support group formation based on Semantic Web techniques (e.g., ontologies). To reduce such limitations, this paper proposes an algorithm based on a CL ontology that matches students’ personality traits to students’ CL roles, which were extracted from CL theories, for creating learning groups.

<table>
<thead>
<tr>
<th>Related Work</th>
<th>Grouping Strategy</th>
<th>Achievement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graf and Bekele (2006)</td>
<td>Mathematical model</td>
<td>- Group diversity (heterogeneous)</td>
</tr>
<tr>
<td>Gogoulou et al. (2007)</td>
<td>Genetic algorithm</td>
<td>- Group diversity (homogeneous, mixed)</td>
</tr>
<tr>
<td>Craig, Horton, and Pitt(2010)</td>
<td>Evolutionary algorithm</td>
<td>- Algorithm flexible in terms of type and number of attributes</td>
</tr>
<tr>
<td>Abnar, Orooji, and Taghiyareh (2012)</td>
<td>Genetic algorithm</td>
<td>- Group diversity (mixed) - Algorithm flexible in terms of type and number of attributes</td>
</tr>
<tr>
<td>Cadavid, Ovalle, and Vicari (2012)</td>
<td>Genetic algorithm</td>
<td>- Group diversity (inter-homogeneous) - Algorithm flexible in terms of type and number of attributes</td>
</tr>
<tr>
<td>Farhangian et al. (2015)</td>
<td>Agent</td>
<td>- Prediction model</td>
</tr>
<tr>
<td>Yan and Chen (2013)</td>
<td>Cluster algorithm</td>
<td>- Group diversity (homogeneous)</td>
</tr>
<tr>
<td>García-Vélez et al. (2016)</td>
<td>Machine learning</td>
<td>- Sporadic online conversation groups</td>
</tr>
<tr>
<td>Konert et al. (2016)</td>
<td>Plugin for Moodle</td>
<td>- Improve group learning performance and commitment</td>
</tr>
</tbody>
</table>

### 8 Conclusion

This paper aimed at presenting a group formation algorithm, named G-FusionPT, to mitigate the lack of studies using learner’s personality traits with pedagogical basis (CL theories) in order to justify the groups’ composition. The algorithm focused on creating groups using the learners’ profile and a CL Ontology. G-FusionPT is a scalable algorithm since it accepts as input a large number of students to form learning groups with different numbers of members.

To evaluate G-FusionPT, we developed a simulation with 300 students based on real-world data and compared the output with two other group formation algorithms: Random and G-Fusion. We ran each algorithm 30 times, varying the group size (three, five, and seven members) and considering the following measures: (1) the average number of learners in well-formed groups, (2) the average number of learners in unfit groups, and (3) the average number of orphan learners. The results demonstrated that G-FusionPT produces the highest average number of learners in well-formed groups, and lowest average number of learners in unfit groups, when compared to the other two group formation algorithms.
Based on our simulation, G-FusionPT has proven to be more effective than G-Fusion and Random algorithms in generating well-formed groups. In this study we focused the group formation process in two characteristics of interest: acquisition of knowledge and mental rigidity level. As a limitation, the Monte Carlo simulation does not reflect other user characteristics that are not incorporated into the assumptions.

As a main theoretical contribution, this research has shown how Affective Collaborative Learning roles (ACL roles), represented in the CL ontology (Figure 2), can be executed by computational systems to support the formation of better learning groups. Our next step is to use the G-FusionPT algorithm to perform controlled experiments with large groups of participants in a real learning context, evaluating the quality of groups created. In addition, we plan to create other group formation algorithms, combining the CL ontology (Figure 2) and learners’ profile (Figure 1), and compare the results with G-FusionPT, regarding formed groups, execution time and efficiency.

References


Brauer, S., & Schmidt, T. C. (2012). Group formation in e-learning-enabled online social networks. In: 15th International Conference on Interactive Collaborative Learning (ICL), (pp. 1–8). IEEE, Villach, Austria. DOI: 10.1109/ICL.2012.6402070 [GS Search]


Cravens-Brown, L. M. (2002). Eysenck and Antisocial Behavior: An Analysis of the Associations Between Personality Styles and Problems with Conduc. 75p. Tese (Dissertation) — Ohio State University, Columbus, Ohio. [GS Search]


