

An artificial intelligence tool for heterogeneous team formation in the classroom

Alberola, J. M., del Val, E., Sanchez-Anguix, V., Palomares, A., Teruel, M. D.

Author post-print (accepted) deposited by Coventry University's Repository

Original citation & hyperlink:

Alberola, JM, del Val, E, Sanchez-Anguix, V, Palomares, A & Teruel, MD 2016, 'An artificial intelligence tool for heterogeneous team formation in the classroom' Knowledge-Based Systems, vol 101, pp. 1-14. DOI: 10.1016/j.knosys.2016.02.010
<https://dx.doi.org/10.1016/j.knosys.2016.02.010>

DOI 10.1016/j.knosys.2016.02.010

ISSN 0950-7051

ESSN 1872-7409

Publisher: Elsevier

NOTICE: this is the author's version of a work that was accepted for publication in Knowledge-Based Systems. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Knowledge-Based Systems VOL 101, (2016) DOI: 10.1016/j.knosys.2016.02.010

© 2016, Elsevier. Licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International
<http://creativecommons.org/licenses/by-nc-nd/4.0/>

Copyright © and Moral Rights are retained by the author(s) and/ or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This item cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder(s). The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

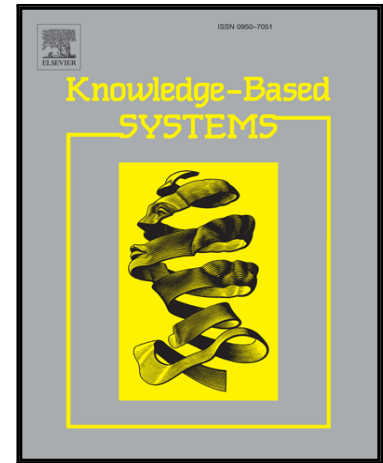
This document is the author's post-print version, incorporating any revisions agreed during the peer-review process. Some differences between the published version and this version may remain and you are advised to consult the published version if you wish to cite from it.

Accepted Manuscript

An artificial intelligence tool for heterogeneous team formation in the classroom

Juan M. Alberola, Elena del Val, Victor Sanchez-Anguix, Alberto Palomares, Maria Dolores Teruel

PII: S0950-7051(16)00096-4
DOI: [10.1016/j.knosys.2016.02.010](https://doi.org/10.1016/j.knosys.2016.02.010)
Reference: KNOSYS 3430



To appear in: *Knowledge-Based Systems*

Received date: 1 September 2015
Revised date: 14 January 2016
Accepted date: 17 February 2016

Please cite this article as: Juan M. Alberola, Elena del Val, Victor Sanchez-Anguix, Alberto Palomares, Maria Dolores Teruel, An artificial intelligence tool for heterogeneous team formation in the classroom, *Knowledge-Based Systems* (2016), doi: [10.1016/j.knosys.2016.02.010](https://doi.org/10.1016/j.knosys.2016.02.010)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

An artificial intelligence tool for heterogeneous team formation in the classroom

Juan M. Alberola^a, Elena del Val^a, Victor Sanchez-Anguix^b, Alberto Palomares^a, Maria Dolores Teruel^a

^a*Universitat Politècnica de València, Cami de Vera s/n, 46022, Valencia, Spain*

^b*Coventry University, Department of Computing, Gulson Road, CV1 2JH*

Abstract

Nowadays, there is increasing interest in the development of teamwork skills in the educational context. This growing interest is motivated by its pedagogical effectiveness and the fact that, in labour contexts, enterprises organize their employees in teams to carry out complex projects. Despite its crucial importance in the classroom and industry, there is a lack of support for the team formation process. Not only do many factors influence team performance, but the problem becomes exponentially costly if teams are to be optimized. In this article, we propose a tool whose aim it is to cover such a gap. It combines artificial intelligence techniques such as coalition structure generation, Bayesian learning, and Belbin's role theory to facilitate the generation of working groups in an educational context. This tool improves current state of the art proposals in three ways: i) it takes into account the feedback of other teammates in order to establish the most predominant role of a student instead of self-perception questionnaires; ii) it handles uncertainty with regard to each student's predominant team role; iii) it is iterative since it considers information from several interactions in order to improve the estimation of role assignments. We tested the performance of the proposed tool in an experiment involving students that took part in three different team activities. The experiments suggest that the proposed tool is able to improve different teamwork aspects such as team dynamics and student satisfaction.

Keywords: team formation, artificial intelligence, Belbin roles, computational intelligence

1. Introduction

In the last few years there has been increasing interest in teamwork skills in the area of Higher Education [1, 2, 3, 4, 5]. Many plans of study and faculties have included general teamwork competence as a part of their educational programs for undergraduate students. The reasons for this inclusion are well-grounded in its pedagogical effectiveness and our current industrial paradigm. Firstly, the area of collaborative learning, supported by computers, promotes collaboration and makes learning more effective [6]. Secondly, the industry has shifted from an individually oriented work environment towards a team-oriented workplace. Nowadays, teams are at the heart of a vast majority of modern companies [7, 8, 9, 10]. Despite the often difficult decision-making tasks involving groups of individuals [11, 12], teams have proven to have an inherent ability to solve the complex problems that are confronted in the current work environment.

Given this context, it is fairly reasonable for Higher Education institutions to place a special emphasis on teamwork skills as a part of every program's learning outcomes. Unfortunately, not every single team is successful in their goals, and many teams fail due to incorrect team dynamics, lack of communication, and interpersonal conflict among team members [13, 12, 14]. Even though some of the aforementioned problems can be alleviated with teamwork experience, these negative factors should be avoided whenever possible as they may generate resentment towards teamwork. Hence, identifying the patterns that drive successful teams and forming work teams according to these patterns become crucial tasks for every organization. Classrooms are not immune to this issue (specially if students are to learn teamwork skills), and unnecessary problems may hinder this learning process.

One of the most important theories regarding successful team dynamics is Belbin's role taxonomy [15]. In this theory, Belbin identifies eight heterogeneous behavioural patterns that are present in many successful teams in the industry: *plant*, *resource investigator*, *coordinator*, *shaper*, *monitor evaluator*, *team worker*, *implementer*, and *finisher*. These behavioural patterns (or roles) should be played by the different team members in order to facilitate successful teamwork. Belbin's taxonomy has given rise to a wide variety of studies showing the theory's strengths and weaknesses [16, 17, 18, 19, 20], it has been applied to a wide variety of domains [21, 22, 23, 24].

As shown by several studies, the classroom environment may benefit from the application of Belbin's theory [25, 26, 27, 28, 29]. One of the reasons for

this successful application in education is the identification of behavioral patterns that are present in many group dynamics. However, there are several problematic circumstances that should be addressed in order to apply Belbin's role taxonomy to the classroom. The first one is that Belbin's roles are classically identified by means of questionnaires (mainly self-perception questionnaires) that are filled out before working with others. However, self-perception results may differ from those patterns shown in a real team environment [30]. Therefore, we believe that a more effective role assignment could be achieved by considering both the information collected before working on a team and the feedback provided from peers after working on a team. The second one is that individuals are not purely described by just a static and strict role. Despite the fact that, due to the individual's personality, one may have a most predominant role, individuals show a rich variety of behavioral patterns depending on circumstances. Firstly, this makes the most prominent role of the individual uncertain, as the individual may show a range behaviors for different scenarios. Secondly, as the individual may show different behavioral patterns, the individual behavior may be best described as a probability distribution over such patterns or roles that he/she plays. The third one is that even a small classroom with 30 students has $\binom{30}{5} = 142506$ different teams of five individuals, and the total number of team configurations for the classroom explodes exponentially with this amount. Finding the best possible configuration poses a computationally expensive problem for it to be solved manually.

In this article, we present a computational tool that attempts to address the aforementioned problems. The tool is based on artificial intelligence (AI) and iterative interactions/feedback. The use of AI techniques allows us to address uncertainty and solve computationally expensive problems. More specifically, the tool makes use of Bayesian learning to tackle uncertainty with regard to students' prominent roles, and the problem of finding optimal teams is treated as a coalitional structure generation problem [31], which is solved by means of linear programming methods. Additionally, the proposed tool is iterative in nature: it proposes team configurations for class task assignments and then it gathers feedback from team members with respect to the roles portrayed by the other teammates. This information is later used to refine future team configurations proposed by the tool.

The remainder of this article is organized as follows. Section 2 describes the main features of the Belbin model. Section 3 presents how the tool would generally work and some implementation details. Section 4 presents an in-

depth explanation of the mechanism used for team formation, which is at the core of our team formation tool. Section 5 analyzes the impact of testing our proposal in a real educational environment. Section 6 shows the most relevant works in the literature with regard to team formation tools. Finally, Section 7 presents some concluding remarks and future work.

2. The Belbin theory

Prior to detailing how the proposed tool was implemented, we believe that it is important for the reader to be familiar with the Belbin theory since it is one of the fundamental pillars of our tool. The Belbin theory [15, 32, 33, 34] provides a thorough of the influence of different types of roles in teamwork. A team role is defined as a behavioral pattern that facilitates the progress of the whole team. Assuming that there would probably be boundless behaviour patterns, Belbin states that the range of behaviours that really influence the performance of a team is limited. In Belbin's model, a role is defined by six factors: personality, mental ability, current values and motivation, field constraints, experience, and role learning [19]. Specifically, Belbin defines the following eight roles:

- **Plant/Creative:** is creative and imaginative. He/she generates ideas and solves difficult problems.
- **Resource investigator:** is outgoing and communicative. He/she explores opportunities and interacts with people outside the team.
- **Co-ordinator:** is mature and confident. He/she has a global view of the project and delegates effectively.
- **Shaper:** is challenging and dynamic. He/she has the drive and courage to overcome obstacles.
- **Monitor evaluator:** is sober, strategic, and discerning. He/she sees all options and judges accurately.
- **Teamworker:** is co-operative, perceptive, and diplomatic. He/she is able to listen and avert friction.
- **Implementer:** is practical, reliable, and efficient. He/she turns ideas into actions and organizes work that needs to be done.

- **Completer finisher:** is painstaking, conscientious, and anxious. He/she searches out errors, polishes, and perfects them.

In later revisions of this theory, a ninth role of *specialist* was introduced for the case when technical expertise is necessary for the performance of certain tasks. Belbin's model has been associated to behaviours and performance. In line with other authors, Belbin has argued that the most successful teams are composed of a balanced combination of the above roles, ideally all of them. In contrast, teams composed of homogeneous roles tend to provide unsatisfactory results.

The Belbin model is traditionally operationalized through the Team Role Self-Perception Inventory, which allows each individual to discover his/her most prominent role based on his/her own judgment. The main disadvantage of this self-perception questionnaire is that individuals may have a preconceived image of themselves, which is diametrically different to the image that is reflected to others [35, 36, 30, 37]. Complementary to this, an Observer Assessment Sheet was also designed to be used by other colleagues who could make an informed judgment based on their knowledge of an individual. However, this questionnaire usually assumed that the observer should know the individual that was being evaluated in depth. This is something that is not always possible to assume in higher education contexts.

3. General tool workflow

In this section, we describe the general workflow of our tool and its most important features. During an academic course, a teacher may carry out several team activities that require the formation of teams. As mentioned above, one of the main problems for teachers is to optimally create teams when there is no previous information about student profiles, and the number of students is high. In the latter scenario, the complexity of determining optimal teams is complex for the teacher due to the exponential nature of the problem [38]. In order to provide support for this team management task, we have developed a software application for teachers that facilitates the costly task of dividing students into optimal or near optimal teams. As a general outline, the application relies on student feedback, coalitional structure generation, and Bayesian learning to form proper distributions of student teams. In the following paragraphs, we will explain how these elements are put together to provide an adequate team formation tool.

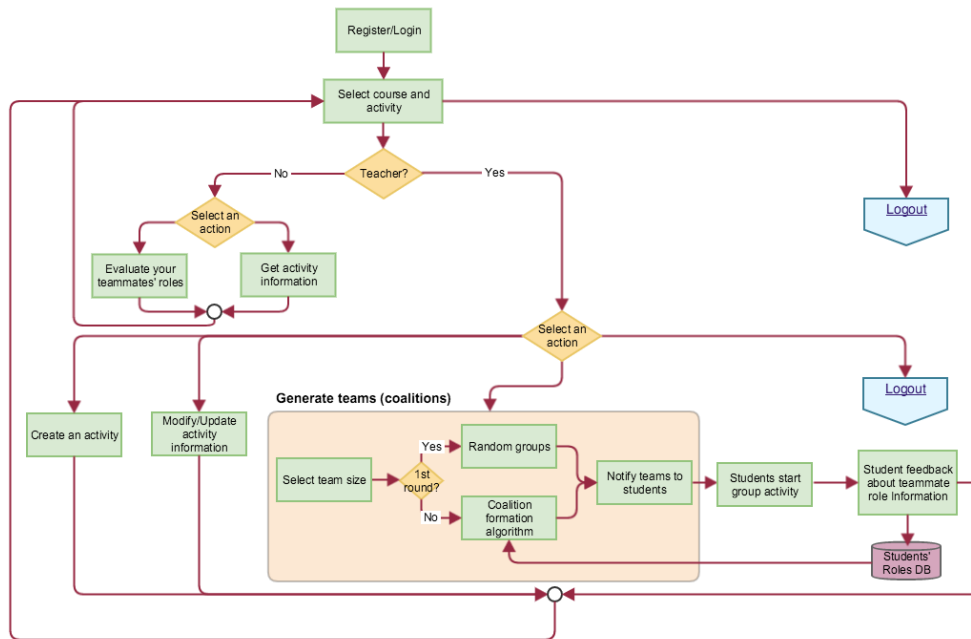


Figure 1: Activity diagram flow for the team formation tool

It should be noted that the tool has been designed to be integrated in web platforms where the actors (i.e., teachers and students) can interact with the system. We have a standalone web application, and we have also worked towards the integration of part of the tool's functionality into Sakai¹, which is the e-learning platform of choice at the Universitat Politècnica de València. The main functionalities of the system can be observed in the UML use case diagram shown in Figure 2. A more detailed diagram flow can be observed in Figure 1.

The initial starting point for the application usually corresponds to the teacher login in the system (see Figure 3, left). The teacher can then create a new team activity. After that, the web application shows the associated modules to the teacher in a pull-down menu. The teacher chooses the module where the team activity will be developed. The teacher can also fill out all

¹<http://sakaiproject.org>

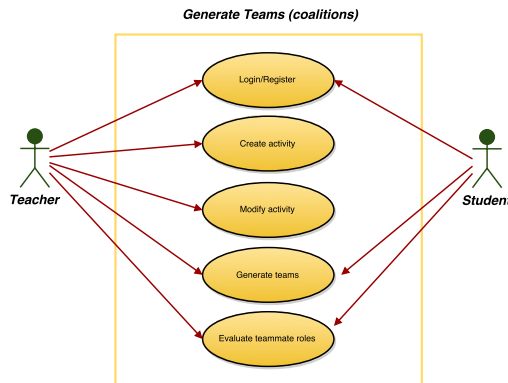


Figure 2: Use cases

the fields associated to the activity (activity description, start date, end date, on-line material for the activity). The teacher should then determine the maximum and minimum size for the student teams.

Taking into account these parameters, the team formation mechanism is ready to generate an automatic proposal of teams. If it is the first time that the tool takes a particular group of students as input, the tool does not have information about the students' roles in the database. Therefore, the first set of teams is generated randomly. Otherwise, the system may employ a more complex team formation algorithm, taking into account the information provided by students in previous team activities. In this article, we employ a team formation algorithm that is based on Belbin's theory, student feedback, coalitional structure generation, and Bayesian learning. This algorithm is described in detail in Section 4. However, we have designed a modular, loosely-coupled web tool where each functionality is provided by a different software module in order to facilitate adaptability to new requirements or modifications. Hence, it is easy to adapt the tool to consider other criteria for group generation (e.g., personality traits, students marks, etc.). Once the teams have been formed, they are sent via email to the teacher, who can in turn notify students about their teammates. The teacher can then publish the team configuration.

Once the activity has been carried out in the classroom, each student receives a notification from the system to provide feedback about his/her teammates according to Belbin's roles. Student feedback is one of the main

Figure 3: (Left) Activity creation form. (Right) Teammate evaluation web form in our standalone application.

aspect of our team formation tool because this information will be used in future team formation tasks. In order to provide feedback, each student should log in the application. Then, the student chooses the module, the activity, and the teammate to be evaluated (see Figure 3, right). In contrast to other Belbin approaches, we have advocated for a simple feedback questionnaire. The reason for this is that students do not get to know their team members well enough during team activities to be able to fill out the Observer Assessment sheet [39] (i.e., a questionnaire used by colleagues of the individual or those who know him/her well). The application shows the description of the roles, and, at that point, the student should assign a role description to each teammate. As the reader can observe, we are using students as classifiers, and the performance of the tool is linked to the ability of humans for such a task. The information is then gathered by the application and stored in a database. Besides the information about the roles of his/her teammates, the student fills out a form with information about the degree of satisfaction with the team and with the methodology. The evaluation of the role of each team member has also been integrated in Sakai (see Figure 4). This integration

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

El meu espai de treball Microsoft Windows Server 2012 2014-Nuevas tecnologías aplicadas al turismo Els meus espais actius

2014-Nuevas tecnologías aplicadas al turismo: Exàmens

Vista prèvia - És un exemple de com veu l'examen un alumne fet

Evaluation of teamwork 1

Pàgina d'index

Part 1 de 1 -

Pregunta 3 de 7

Identify which of the following roles best fit each of your partners:

Creative: is creative and imaginative. He/she generate ideas and solve difficult problems.

Resource Investigator: is outgoing and communicative. He/she explores opportunities and interacts with people outside the team.

Co-ordinator: is mature and confident. He/she a global view of the project and delegates effectively.

Shaper: is challenging and dynamic. He/she has the drive and courage to overcome obstacles.

Monitor evaluator: is sober, strategic and discerning. He/she sees all options and judges accurately

Teamworker: is co-operative, perceptive and diplomatic. He/she is able to listen and avert friction.

Implementer: is practical, reliable and efficient. He/she turns ideas into actions and organises work that needs to be done.

Completer finisher: is painstaking, conscientious and anxious. He/she searches out errors, polishes and perfects it.

	Creative	Resource Investigator	Co-ordinator	Shaper	Monitor evaluator	Teamworker	Implementer	Completer finisher
Sergio	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Claudia	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Alba	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Selena	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Donka	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure 4: Team evaluation web form in Sakai

facilitates the provision of the data by the students since the feedback form is available in the module of the course. As new team activities are carried out, new information is gathered and accumulated so that this information can be used in future team formation tasks. The idea behind this mechanism is that as the system gets more information, the system would have more evidence about the predominant roles of each student (i.e., the most often observed behavioral patterns) and therefore better team allocations can be provided.

4. Team formation mechanism

At this point, we have described the general use of the tool. In this section, we describe our policy for dividing students into teams. This policy relies on student feedback, coalitional structure generation, and Bayesian learning to form proper distributions of student teams. First, we describe how dividing students into optimal teams is equivalent to a coalition structure generation

problem. Then, we describe how Bayesian learning is employed to update the information of the classroom and to determine the most relevant role of each student.

4.1. Student team formation as a coalition structure generation problem

The Coalition Structure Generation problem refers to partitioning the components of a set into exhaustive and disjoint coalitions so that the global benefits of the system are optimized. In our problem, the components of the set are the students that take part in a classroom team activity:

Definition 1. Let $A = \{a_1, \dots, a_n\}$ be a set of students, and let $R = \{r_1, r_2, \dots, r_m\}$ be the set of roles that a student may play (in our case, Belbin's roles), and let $role_i$ denote the most predominant role of a_i . Let $T_j \subseteq A$ be a subset of A called team.

The value of a team T_j is given by a characteristic function $v(T_j)$. A characteristic function $v(T_j) : 2^A \rightarrow \mathbb{R}$ assigns a real-valued payoff to each team T_j . The value of a team $v(T_j)$ is calculated on the basis of the most predominant role that each student $a_i \in T_j$ has ($role_i$). Let k denote the size of the team T_j and $\pi_j = \{role_1, \dots, role_k\}$ with $\forall r'_i \in R$ be a vector with the most predominant role of each team member of T_j . In that case, $v(T_j) = v(\pi_j)$.

Unfortunately, it is not possible to know the main role of each team member π_j exactly and therefore $v(\pi_j)$ cannot be calculated precisely. However, it is possible for us to calculate an estimation of the value of the team $v(T_j)$ given the history of evaluations received for each student H that is gathered and accumulated after each team activity. Let $\pi'_j = \{role_1 = r'_1, \dots, role_k = r'_k\}$ be a vector containing a set of hypotheses for the most predominant role of each team member, and let Π be the set of all possible vectors of hypotheses for predominant roles of T_j . We calculate the expected value of a team given the history of evaluations as:

$$\hat{v}(T_j|H) = \sum_{\pi' \in \Pi} p(\pi'|H) \times v(\pi') = \sum_{\pi' \in \Pi} \left(v(\pi') \times \prod_{a_i \in T_j} p(role_i = r'_i|H) \right) \quad (1)$$

where $p(\pi'|H)$ represents the probability for π' to be the real role distribution in T_j given the history of evaluations H . Each $p(\pi'|H)$ can be split into its

$p(\text{role}_i = r'_i | H)$ since we assume that the role of each student is conditionally independent of other students' roles given the history of evaluations.

Conforming to different studies [40], the team should take advantage of having a balanced distribution of roles (i.e., one person per role) and is expected to obtain better results than a homogeneous team. In the coalition generation problem presented in this article, the overall effectiveness of each team $v(\pi')$ is measured numerically depending on the distribution of roles in the team as follows:

$$v(\pi') = \frac{1}{2^{D(\pi')}}, \quad (2)$$

where $D(\pi')$ calculates the number of roles that are repeated in π' .

This function exponentially penalizes repeated roles in a team. For instance, a team that consists of four students all of whom have distinct roles will have a maximum efficiency of 1 (the best team configuration); a team of four where there are two students with the same role will have an efficiency of 0.5 (this value tries to reflect that temporary arguments might arise in the team where students with equal roles appear); a team where there are three students with the same role will have an efficiency of 0.25 (this value reflects that there is a higher probability that many problems will arise on the team due to the high number of students with a similar profile and the efficiency of the team would be affected); and finally, a team where the four students have the same dominant role will have an efficiency of 0.125.

Once we have defined how a team is created and its associated value, we define the concept of *team structure*.

Definition 2. A team structure $S = \{T_1, T_2, \dots, T_k\}$ is a partition of teams such that $\forall i, j (i \neq j), T_j \cap T_i = \emptyset, \bigcup_{\forall T_j \in S} T_j = A$.

The value of a team structure is denoted by $v(S)$, where $v(S)$ is an evaluation function for the team structure. In this work, we consider that the quality of each team is independent of other teams. Therefore, we can calculate the value of the team structure as $v(S) = \sum_{T_j \in S} v(T_j)$.

The goal of the application is to determine an optimal team structure for the classroom $\underset{S \in 2^A}{\operatorname{argmax}} v(S)$. Taking into account Equation 1 to calculate the expected value of a given team, our team formation problem requires

that the following expression be solved at each iteration:

$$\operatorname{argmax}_{S \in 2^A} \sum_{T \in S} \hat{v}(T|H) \quad (3)$$

It turns out that partitioning a set of students into disjoint teams while optimizing a social welfare function corresponds to the formalization of coalition structure generation problems. In order to solve this problem, we formally define the coalition structure generation problem as a linear programming problem [41] and solve it with the commercial software *ILOG CPLEX 12.5*².

4.2. Estimating students' roles via Bayesian learning

After every activity, each student evaluates his/her peers by stating the most predominant role of each of his/her teammates. Therefore, new information about the most predominant role of each student is available and the history of evaluations, H , grows. Hence, at each iteration, the application updates information regarding the probability for a student a_i to have r'_i as his/her most relevant role given the evaluation history $p(\text{role}_i = r'_i|H)$. We employ Bayesian learning for this matter:

$$p(\text{role}_i = r'_i|H) = \frac{p(H|\text{role}_i = r'_i) \times p(\text{role}_i = r'_i)}{\sum_{r \in R} p(H|\text{role}_i = r) \times p(\text{role}_i = r)} \quad (4)$$

where $p(H|\text{role}_i = r'_i)$ is the likelihood function and $p(\text{role}_i = r'_i)$ is the prior probability for the hypothesis. For the likelihood function, we can calculate it as $p(H|\text{role}_i = r'_i) = \frac{\#\{r'_i \in H_i\}}{|H_i|}$, where H_i denotes the peer evaluations about agent a_i and $\#\{r'_i \in H_i\}$ indicates the number of times that r'_i appears as evaluation in H_i . As for the prior probability, we calculate it as $p(\text{role}_i = r'_i) = \frac{\#\{r'_i \in H\}}{|H|}$. Laplace smoothing [42] is employed to ensure that the likelihood for each role hypothesis can be calculated in the first iterations.

This way, we build a probability distribution for each student over the roles that he/she plays. This mechanism allows us to model the fact that individuals are complex, and they may show a variety of behavioral patterns (although one of them may be the most predominant one). In Equation 1 we use this information to estimate the expected value for a team given the probability distribution of students over behavioral patterns.

²<http://www.ibm.com/software/commerce/optimization/cplex-optimizer/>

5. Experiments

This tool was validated in an experimental setting involving real students. More specifically, the experiment was carried out in two consecutive modules in the *Tourism Degree Program of the Universitat Politècnica de València* during the academic year 2013/2014. The first module is held during the first semester, whereas the second is held during the second semester. Both modules are compulsory for students, and they are placed in the same academic course. Therefore, students attending the first module usually also attend the second module.

There may be some students that only participate in one of the two modules (i.e., the first or the second module). However, one of our goals is to check the robustness of the proposed tool in a realistic setting where the tool has less information regarding some of the students. The reason for selecting these two modules is that they have been classically taught using cooperative learning methods.

5.1. Experimental setting

The experiment was divided into three team activities that were split between the two modules. The first activity performed in the first module, while the second and third team activities performed in the second module. Prior to the start of each team activity, the tool proposes a set of teams ranging from 4 to 6 members based on the available information. For the preparation of the first team activity, since there is no information about students, the students are randomly placed into teams. We advocated for the use of random teams instead of letting the students decide their own teams.

A total of 77 students participated at some point in the experiment. Of that set of students, 60 students participated in the first team activity and 50 students participated in the second and third team activities. A total of 34 students participated in all of the team activities, while 19 additional students participated in the second and third team activities. It should be noted that the first team activity acts as the benchmark for team performance, and the second and third team activities reflect the effect of the proposed tool on team performance. Even in the case of the second activity, where 19 additional students took part in the activity, a majority of the students (i.e., 34 students) already had peer evaluation from the first activity.

Apart from the peer evaluation that is carried out after each team activity, we devised a questionnaire that contains questions about different teamwork aspects such as communication, coordination, team dynamics, etc. These questionnaires were answered by students at the end of each team activity, and, in general, they could reflect different levels of satisfaction using Likert scales. The questions have been organized according to two main categories: team dynamics (i.e., coordination, cooperation, work balance, team member behaviour, decision making, etc.) and student satisfaction. The specific questions can be found in Table 1.

5.2. Results

In order to analyze the performance of the proposed tool regarding satisfaction at the team and the individual levels, we compared the questionnaires from the first team activity versus the questionnaires obtained after the second and the third team activities. To analyze data from Likert scales, we first combined categories into binary categories (i.e., disagree or indifferent/agree) since some of the options did not have enough samples to generalize (i.e., less than 5 counts).

When combined into binary categories, the result for each question are two 2x2 contingency tables, one for the comparison between the first team activity and the second team activity and one for the comparison between the first team activity and the third team activity. There are two significant problems when studying team formation in the classroom. On the one hand, many classes are not composed of a large number of students, and, therefore, the number of samples per experiment tends to be low and it is difficult to include more samples. In this sense, the setting resembles that of the life and medical sciences. On the other hand, it has been reported in the literature that, when the studied variable is discrete (as in our contingency tables), classic calculations of the p-value do not represent its classic meaning [43, 44, 45, 46, 47, 48]. Due to the discrete nature of the variable, only one set of p-values is possible and the method tends to be excessively conservative and far from the meaning of classic p-values. For these scenarios, researchers propose the calculation of the mid p-value, whose type I error rate is closer to the nominal level. We employed a one-tailed test ($\alpha = 0.05$) to test if the ratio of positive answers in the second activity is greater than the ratio of positive answers in the initial activity [49, 50]. For these tests, we calculated the mid p-value since its type I error rate is closer to the nominal level. Similarly, we carried out another one-sided proportion test ($\alpha = 0.05$) to test

Table 1: The questionnaire handed to students after each team activity

Team dynamics	
Coordination & Cooperation	
Q1. Our team has been well-coordinated	Likert scale
Q2. Our team created a cooperative environment	Likert scale
Q3. We established specific norms and tasks for everyone	Likert scale
Q4. Each member did his/her share of work	Likert scale
Q5. Our team spent its time well	Likert scale
Team member behaviour	
Q6. The attitudes of team members were positive	Likert scale
Q7. Team members treat each other in a respectful way	Likert scale
Decision making	
Q8. Team members accepted suggestions in a positive way	Likert scale
Q9. Our team listened to the opinions of all the members	Likert scale
Q10. It was difficult for my team to come to a decision	Likert scale
Student satisfaction	
Q11. How was your experience on this team?	Likert scale
Q12. I felt comfortable working on my team	Likert scale
Q13. I felt that I contributed to this team	Likert scale
Q14. Our project is acceptable to everyone on the team	Likert scale
Q15. Evaluate your team with a grade from 0 to 10	0 to 10 grade

if the proportion of positive answers in the final activity is greater than the proportion of positive answers in the initial activity. In order to account for possible type I errors, we adjusted the mid p-values obtained using Hommel's method.

As for numeric answers (Q15), a one sided Mann Whitney statistical test ($\alpha = 0.05$) was employed to assess statistical differences between the null hypothesis and alternatives, with the null hypothesis representing the same underlying distribution, and the alternative hypothesis representing an underlying distribution whose values are larger than those presented by the other distribution. In order to account for possible type I errors, we adjusted obtained p-values by using Hommel's method.

The results for the experiment can be viewed in Table 2. The table shows (in bold font and green shadow cell) those scenarios where the alternative hypothesis was accepted with 95% confidence and it held significantly better results than those gathered in the initial activity. A blue shadow cell is used to highlight those values that are close to 95% confidence. It should be highlighted that there was not a single question where the proposed tool obtained worse results than the initial team activity. The proposed tool always offered answers that were at least as good as the initial setting, and, in many cases, the responses were statistically more satisfactory. Out of the 15 answers gathered, 9 questions obtain statistically better results than the initial activity by the last iteration of our proposed method (activity 3). One additional question obtained results that are close to being statistically significant. In the first iteration of our proposed tool (activity 2), 8 questions obtained statistically better results than the initial activity, and one additional question was close to being statistically significant. Hence, even in intermediate iterations, the proposed tool is capable of obtaining better responses from students. We comment on these results in more detail below.

5.2.1. Team dynamics

The results for the questions related to team dynamics supported the good performance of our proposed tool. Of the 10 questions concerning team dynamics, 5 showed statistically better answers in the third activity (Q1, Q2, Q3, Q5, Q6) than those obtained in the initial activity, and 1 extra question was close to obtaining statistically better answers in the third activity (Q4). Hence, at the end of the experiment, the proposed tool was capable of improving 5 out of the 10 team dynamics without decreasing the quality of the other 5 team dynamics. In the second activity (which

Table 2: The results of the questionnaire for the three team activities. D: disagree, N: neutral, A: agree

		Act. 1	Act. 2	Act. 3
Team dynamics	Q1	D:12, A:36	D:4, A:40	D:3, A:37
	Q2	D:15, A:33	D:2, A:42	D:4, A:36
	Q3	D:11, A:37	D:2, A:42	D:3, A:37
	Q4	D:16, A:32	D:6, A:38	D:8, A:32
	Q5	D:11, A:37	D:5, A:39	D:3, A:37
	Q6	D:16, A:32	D:8, A:36	D:4, A:36
	Q7	D:8, A:40	D:5, A:39	D:4, A:36
	Q8	D:7, A:41	D:6, A:38	D:3, A:37
	Q9	D:25, A:23	D:25, A:19	D:26, A:14
	Q10	D:35, A:13	D:33, A:11	D:27, A:13
Student satisfaction	Q11	D:7, A:41	D:0, A:44	D:1, A:39
	Q12	D:18, A:30	D:5, A:39	D:5, A:35
	Q13	D:7, A:41	D:2, A:42	D:3, A:37
	Q14	D:10, A:38	D:1, A:43	D:3, A:37
	Q15	8.05	8.22	8.41

represents an intermediate iteration of the proposed tool), it was also capable of improving 4 out of the 10 team dynamics (Q1, Q2, Q3, Q4) without decreasing the quality of the other 5 team dynamics. Additionally, another question was close to achieving statistical significance (Q5,Q6).

The first set of questions addresses issues that are related to cooperation and coordination (Q1, Q2, Q3, Q4, Q5), which are critical team performance factors [51, 52]. First of all, a higher percentage of students perceived a cooperative environment in those teams proposed by the tool (Q2). On the one hand, an adjusted mid p-value of 0.008 and the ratio of students that answered this question positively increased by 26 (from 69% to 95%) if we compare the first and the second team activities. On the other hand, an adjusted mid p-value of 0.008 and an increase of 21 in the ratio of students that answered this question positively (from 69% to 90%) was obtained when comparing the first and the third team activities. This shared sense of cooperation was also accompanied by a clearer perception of good coordination among team members (Q1). In this case, if we compare the first and the second team activities, the adjusted mid p-value was 0.02 with an increase of 15% in the ratio of students that that answered this question positively (from 75% to 90%). When compared with the first team activity, the third team activity obtained an adjusted mid p-value of 0.02 with an increase of 17% in the ratio of students that answered this question positively (from 75% to 93%). A detailed view on how students answered these two questions is shown in Fig. 5. As can be observed, the ratio of students that answered these both questions positively for the second and third team activities is larger than the ratio of students that answered these both questions positively on the first team activity.

It is likely that good coordination and cooperation were possible due to the fact that those teams proposed by the tool were able to more frequently agree on clear team norms and commitments (Q3, Q4), and time was managed in a more efficient way (Q5). In the case of Q3, the adjusted mid p-values obtained when comparing the second and the third team activities with the first team activity were 0.01 and 0.02, respectively, whereas there was an increase of 18 and 16 in the ratio of students that answered this question positively (from 77% to 95% and 93%, respectively). The results for question Q4 show an adjusted mid p-value of 0.02 when comparing the first and the second team activities with an increase of 20 in the ratio of students that answered this question positively (from 66% to 86%). When comparing the first and the third activities, the adjusted p-value was found to be 0.08

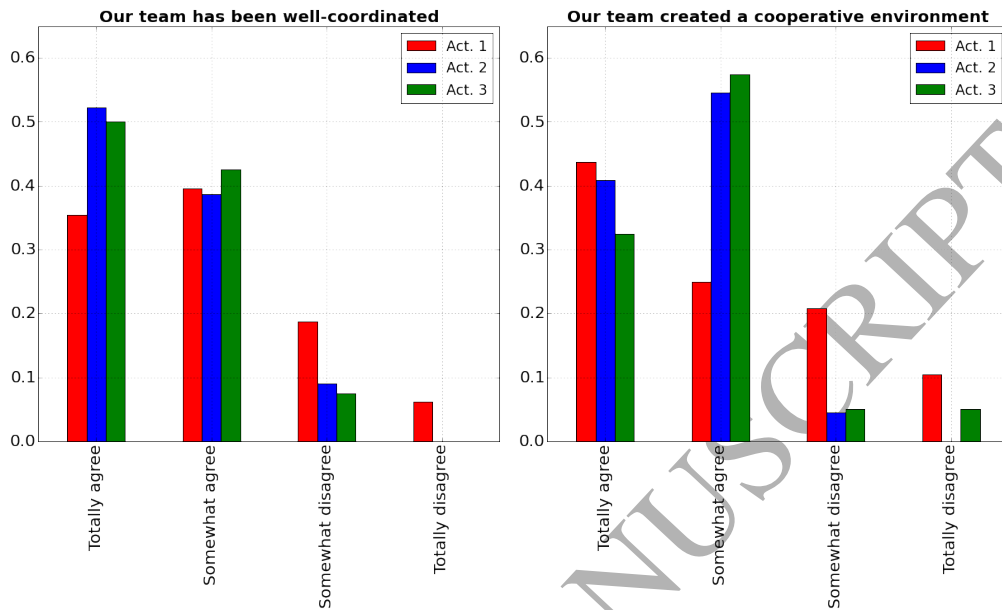


Figure 5: Detailed answers for Q1 (left), and Q2 (right)

and we cannot conclude that students considered that all of the teammates did their share of the work. Despite this, the mid p-value is very close to statistical significance. Hence, we consider that the result for this question is partially supported. Nevertheless, we observed an increase of 14 in the ratio of students that answered this question positively (from 66% to 80%) when comparing the first and the third team activities. For Q5, we could not find a significant statistical difference between the first and the second team activities, with an adjusted mid p-value of 0.07. However, we did observe an increase in the ratio of students that answered this question positively (from 77% to 88%). We observed statistical differences between the third and the first team activities. In that case, the adjusted p-value was 0.04 and there was an increase of 16 in the ratio of students that answered this question positively (from 77% to 93%). Hence, taking into account all of the previous findings, it can be observed that the proposed tool is able to form teams that attain higher levels of coordination and cooperation (i.e., Q1, Q2, Q3, Q5, and partially Q4).

The second set of questions involves team member behaviours within

the team (Q6, Q7). As pointed out by the results, a significant and higher percentage of students perceived a positive attitude in team members when the teams were proposed by our tool (Q6) during the third team activity. The adjusted mid p-value for this comparison was found to be 0.009 with an increase of 24 in the ratio of students that answered this question positively (from 66% to 90%). This positive attitude may be a direct consequence of (or be reinforced by) the general feeling of cooperation and coordination (Q1, Q2). The comparison between the first and the second team activities obtained an adjusted mid p-value of 0.05, which only partially supports the finding, and there was an increase of 16 in the ratio of students that answered this question positively (from 66% to 82%). Even though the tool was not able to generate teams that perceived a more positive attitude in teammates in the second activity, after the third iteration, the algorithm was capable of providing a positive environment with the teams that were generated. The reason for this may be explained by the fact that the algorithm has more information with which to provide appropriate team allocations.

Even though team members had a more positive attitude, we could not find any significant statistical differences regarding how students treated their peers throughout the experiment (Q7). The comparison between the first and the second activities obtained an adjusted p-value of 0.24, and the comparison between the first and the third activities also obtained an adjusted p-value of 0.24 too. However, there was an increase in the ratio of students that answered this question positively in the second and third activities with respect to the first team activity (from 80% to 88% and 90%). In all of the cases, the students were generally respectful with others and there was no serious incidents during the experiment.

When it comes to decision-making aspects (Q8, Q9 and Q10), the experiments showed no support for better decision making. Decision-making processes are difficult, especially when they involve groups of individuals. Despite the fact that Belbin's theory has been reported to improve team performance in complex decision-making tasks like management games [53], we could not find significant improvements over the initial team configuration. In contrast, to other works where Belbin roles have contributed significantly to decision-making tasks [53], we believe that the main difference between our present scenario and other works resides in the fact that decision making is not the core activity of the projects proposed to students. Naturally, there is always a decision-making component that is inherent to every team activity, but it is perhaps not as strong as in management games. Consequently,

it is our belief that differences in decision-making processes are more difficult to observe.

5.2.2. Student satisfaction

As Table 2 shows, in the last iteration, the proposed tool is able to obtain statistically better answers to 4 of the 5 questions concerning student satisfaction. More specifically, a statistically higher percentage of students evaluated the team experience as positive and felt more comfortable working on the teams proposed by our team formation tool (Q11, Q12), even in the early stages. In the case of Q11, we observed adjusted mid p-values of 0.008 and 0.02, and an increase of 16 and 16 in the ratio of students that perceived positive aspects in the team experience (from 66% to 82%) when comparing the first activity with the second and the third team activities, respectively. For Q12, it was observed that adjusted mid p-values strongly supported the use of our tool in the second and third activities with respect to the classic team formation mechanism (adjusted mid p-values of 0.004 and 0.004, respectively). We also observed a dramatic increase in the ratio of students that answered this question positively (from 62% to 88% and 87%). A more detailed view of student satisfaction with the team activities can be found in Figure 6. As in other figures, the reader can observe how the number of students who answered this question positively is larger for the second and third team activities than for the first team activity.

This is also partially reflected in the evaluations given by students for their respective teams (Q15). Also note that since the tool incorporates more feedback from the students with each activity, the average team evaluation gets higher. Therefore, in the third team activity, the result obtained for Q15 is statistically significant (p-value of 0.04) and better than that for the first team activity (an increase of 0.4 on a scale of 0 to 10).

We cannot confirm that a higher percentage of students perceived that they contributed to the team (Q13). Adjusted mid p-values were not close to the α value of significance (0.12 and 0.16 for the second and third activities, respectively) despite the fact that we found increases in the ratios of students that answered this question positively (from 85% to 95% and 93%). In none of the activities did the students feel that they had not contributed to the team. This result should be treated carefully since students may be biased when self-assessing their own contributions [54].

Not only were the students more satisfied with the team, but a higher percentage of students perceived that the work carried out in the group was

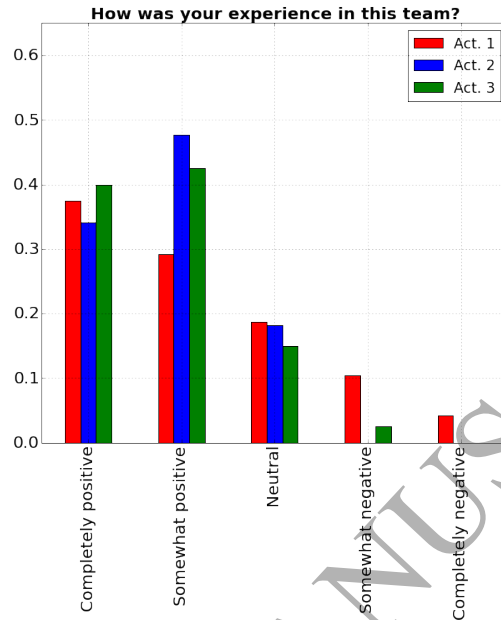


Figure 6: Detailed answer for Q11

satisfying for other members on the team (Q14). In this case, the comparison between the first and the second team activities obtained an adjusted mid p-value of 0.006 with an increase of 19 in the ratio of students that answered this question positively (from 79% to 98%). The comparison between the first and the third team activities also offered positive results with an adjusted mid p-value of 0.04 and an increase of 14 in the ratio of students that answered this question positively (from 79% to 93%). Therefore, if previous findings are considered in this experiment, the students perceived a higher degree of individual and peer satisfaction.

5.3. Other findings

The data that we obtained from the experiments allowed us to explore other issues which are indirectly related to the proposed tool's performance. We analyzed the role distribution in the classroom based on the feedback received from teammates, and considering the most voted role as the most primary role of a student and the second most voted role as the secondary role, . We would like to clarify that this role distribution is by no means

generalizable because students from different academic backgrounds (e.g., art, engineering, etc.) may be different and thus play different roles on teams. Nevertheless, we believe that this role distribution may offer insights for future research.

Figure 7 shows the distribution of roles for the students that participated in the study. The results suggest that students perceived the plant, monitor, team worker, and resource investigator to be the most prominent roles in the classroom, accounting for almost 77% of the students. This first look at the role distribution clearly indicates that roles are not equally distributed among students. Thinking roles (i.e., monitor and plant) accounted for 49.4% of the student population in our experiment, social roles (i.e., coordinator, investigator, and teamworker) accounted for 35.1% of the students, and action roles (i.e., shaper, finisher, and implementer) only accounted for 15.6% of our student population.

As stated by Belbin's theory [15], individuals do not play just a single role on a team; they play several roles as needed. A closer look at the second most prominent or secondary role indicates that one of the most frequent secondary roles for students was implementer, which is an action-oriented role. Hence, action-oriented roles may not be as misrepresented as one might initially think. Given the way that our tool computes its optimality function, which takes into account uncertainty and considers as many role allocations as possible (see Equation 1), the proposed algorithm is able to cope with situations where individuals are not described by single roles.

For the vote distribution for individual roles, we observed that the most prominent role for each student on average received 43% of the votes, while the second most prominent role received 24% of the votes. These results are certainly far from a random vote, where each role would receive 12.5% of the votes. This means that there is a certain consensus among students with regard to peer roles: the most prominent role has almost four times more support than a randomly selected role, whereas the secondary role has almost twice the support than random voting would have.

In previous simulations that we carried out to assess the viability of the proposed tool [55], the tool was even robust to human voting with a high level of noise. As stated above, there is a certain level of consensus with regard to the most prominent roles for students. Therefore, in this situation, it is only logical for the proposed tool to be able to detect the most probable roles for students and act accordingly.

We were also able to analyze this behaviour in our data. Of the 34 stu-

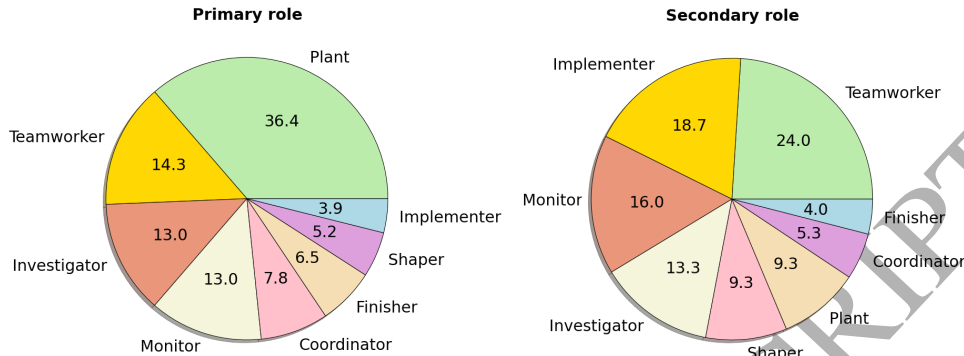


Figure 7: Distribution of the primary (left) and secondary (right) roles of the students who took part in the experiment

dents that participated in all of the team activities, 11 students had their most prominent role converge after the first team activity's feedback, and it was maintained throughout the experiment. An additional set of 12 students had their most prominent role converge after the feedback from the second team activity. Therefore, almost 68% of the students had their most prominent role converge in just two team activities. We claim that this is a positive result, especially if we consider that the tool has been designed to share information across different modules. Therefore, the most prominent role of students is available after a few activities.

6. Related work

The study of the composition and performance of collaborative teams is a topic that has attracted the attention of many researchers for years [56, 57, 58, 59, 60, 61, 62]. Collaborative learning is widely accepted as a particularly important factor that enhances the learning process in educational contexts [63, 64, 65, 66, 67]. However, there are not very many studies that focus on the key issue of team formation, which is a task that requires a considerable amount of time to compose well-balanced teams. More specifically, when there is a large number of students and different grouping criteria, the task of forming collaborative learning teams to promote successful outputs is considered an exponential problem. In line with this, several computational-based approaches have been proposed in the last few years to provide support

to automatically, efficiently, and effectively deal with this goal.

The majority of these proposals are oriented to creating heterogeneous teams since there is an extended view in the literature of a direct relationship between the performance of a team and the heterogeneity of its team members. As an example, Graf et al. [68] present an Ant Colony Optimization approach that provides heterogeneous groups based on the personal traits of students. The groups are made up of by four students with low, average, and high student scores. The algorithm tries to maximize the diversity of the group while keeping a similar degree of heterogeneity in all of the groups.

Genetic or evolutionary algorithms are also commonly used to solve the problem of forming collaborative learning teams. Wang et al. [69] present an approach for automatic team formation based on thinking styles [70] to determine the features of the students. The algorithm translates the features of the students into points in a two-dimensional space which are then classified into categories. The algorithm uses a genetic algorithm to create the optimal group formation based on the categories of the students.

Bergey and King [71] present a decision-support tool to create heterogeneous student teams based on genetic algorithm with grouping criteria consists of characteristics such as work experience, personality, demographics, undergraduate degree, and academic performance. This tool was applied in university contexts, and the performance of these optimally balanced teams was better than those created manually.

Lin et al. [72] present a system that assists instructors in forming collaborative learning groups. The algorithm considers two criteria: information about the understanding levels, and the interests of the students. Particle swarm optimization is used in the group composition algorithm to deal with the complexity of the problem. Even though this proposal presents an algorithm that finds solutions within a reasonable time, the validation of this proposal only focuses on the computational side. Thus, the grouping criteria has not been tested in a real environment in order to measure whether or not the performance of the teams improved.

Cavanaugh et al. [73] present a web-based system to create teams that takes into account characteristics such as gender, skills, and student schedules. Moreno et al. [74] provide an approach with underlying genetic algorithms for inter-homogeneous and intra-heterogeneous team formation. In contrast with the above approaches, this proposal does not limit team formation to a specific number of student characteristics.

Another type of algorithm is used by Christodoulopoulos et al. [75], who

presents a web-based group formation tool that facilitates the creation of homogeneous and heterogeneous groups based on the learning styles of the students. This tool allows the instructor to manually modify the groups and allows the students to negotiate the grouping. The creation of homogeneous groups is based on a Fuzzy C-Means algorithm, and the creation of heterogeneous groups is based on a random selection algorithm. The tool also provides an option to negotiate the proposed teams with the students. However, this negotiation consists of direct interaction with the teacher.

The problem of team formation is also present in the context of human resource management. As an example, Wi et al. [76] present a framework to deal with team formation in R&D-oriented institutes. The authors propose a genetic algorithm that uses a fuzzy model to take into account information about the candidates' knowledge and expertise about certain topics related to a project. The algorithm also takes into account information about the position of the candidates in a social network in order to determine their suitability for project management positions.

As can be observed, there is a large number of computational-based approaches for automatically forming collaborative learning teams that are focused on the individual characteristics of the students (e.g. personality, learning style, learning achievement, or thinking style). These characteristics motivate and enhance the capability of each individual to occupy different team roles, considering a role as the set of goal-directed behaviours taken on by a person within a specific situation [77, 19, 78].

Even though team roles are considered to be a critical part of making effective teams [79, 32, 80, 81, 69] and there are several taxonomies of team roles in the literature [82, 32, 34, 83, 84, 85], there are very few computational-based approaches that focus on this topic.

In [86] and [87], the authors propose an approach that uses an underlying model proposed by Mumma [88], which is based on team roles. This approach considers team formation to be a constraint satisfaction problem. Even though the formation tool is applied in a real context, the experiments focus only on the performance of the tool for finding the groups. Thus, the real impact of team performance is not analyzed.

In the literature, one of the most widely used frameworks for team formation is the Belbin model [32, 34]. This model has been widely applied in different contexts such as organizations, firms, and executive education programs [21, 53, 89, 90, 19, 91]. The Team Role Self-Perception Inventory, which is filled out by each student, is the most frequently used tool to collect

the information of each individual [21, 16, 23, 30, 92]. However, there are some studies that also consider the Observer Assessment Sheet [30, 93].

Despite the popularity of the Belbin model, there are only a few computational-based approaches that focus on team formation based on this model. One of these approaches was proposed by Yannibelli and Amandi [29, 94]. It is based on hybrid evolutionary algorithms to find the most suitable collaborative learning teams in the context of software engineering modules. The grouping criterion is defined on the basis of the Belbin model and uses the Self-Perception Inventory. Thus, the algorithm balances the roles and the number of members. The validation of this proposal is based on computational experiments. Therefore, although the performance of the algorithms is promising, the effectiveness of the approach in a real context was not tested.

Fares and Costaguta [95] present a multi-agent system that promotes team roles in groups in order to achieve successful learning. This proposal uses interface agents to diagnose the state of the collaboration of each individual within the group, in order to determine each of the roles proposed by Belbin. Similarly to the previous approach, this proposal was not validated in a real context.

In comparison with the above approaches, our proposal presents three different contributions that go a step further into the state of the art. First, our proposal contributes to the development of computational solutions for the team formation problem, which is based Belbin's model. In [96], we evaluated our proposal with synthetic data in order to test the feasibility of the tool. However, to validate our proposal, we decided to test the performance of the tool in a real experiment involving students, which corresponds to the results presented in this article. Even though the time performance of the algorithm is important, we do not consider the computational cost in this domain to be critical because the task of team formation is only carried out a few times each academic year. Thus, our validation focuses on showing the quality of the team formation model in a real scenario and testing the impact of applying the Belbin model for team formation. Actually, as observed in the real experiments, the majority of the students have a prominent role that converges in two team activities. This conclusion is very similar to the convergence rate obtained with synthetic data.

The second key contribution of our proposal is related to the information used as the input for the team formation process. In all of the previous approaches, the information of each student is collected prior to working on a team. More specifically, the Self-Perception Inventory (or other psychologi-

cal tests used in different models) are used by other approaches to establish the role of each individual. This role assignment is not expected to change and, therefore, the composition of the team would be similar if several iterations were carried out with the same population of students. However, the behaviour of an individual in a real team environment may be different than one's own perception since moods, emotions, self-esteem and trust can change during the teamwork experience [30, 97, 98, 99]. In our proposal, each student is evaluated by his/her team partners after the development of some joint project or activity. Thus, the role profile considers the aggregated opinion of the rest of the students who worked with him/her. Since the evaluation is carried out after exhibiting the behaviour in a real situation, the role assignment may be more reliable and precise. What is more, the collected information increases as the number of teamwork projects increases.

Although some of the above approaches include student feedback as a measure of satisfaction, student feedback is not used to improve team formation. Similarly, non-computational studies that apply the Observer Assessment Sheet do not consider that the information retrieved may be different depending on the team activity and that individuals are required to know each other well (which is not always feasible in higher education contexts). This led us to the third key contribution of our proposal, which is related to how roles are assigned to individuals. Since roles are established by more than just an initial questionnaire, the role assigned to each individual is not necessarily limited to a single strict role. In contrast to other approaches, we use probabilities to establish how roles are played by each student. Therefore, we do not establish that an individual strictly plays a predominant role, but we establish that an individual plays each of the eight roles with a specific associated probability. This allows us to account for the fact that humans display a variety of behaviors in real situations.

It must be noted that our proposal does not need previous information about the abilities, attributes, or roles played by the students. The information used to characterize the role of each individual is gathered and grows each time a different project or activity is carried out. To our knowledge, our work is the first one that provides this aspect; this improves the information that is collected from the students and establishes a more reliable role assignment since it considers the opinion of other members instead of self-perception.

7. Conclusions

In this article, we have presented an artificial intelligence tool for team formation in the classroom. The tool forms teams based on Belbin's role theory, which identifies eight different behavioural patterns (i.e., roles) that are found in successful teams. As proposed by the theory, teams are formed with a heterogeneous configuration of roles. These roles are inferred from feedback that is received from students with respect to their teammates after each team activity. Artificial intelligence allows us to handle two important issues for the tool: (i) it can handle uncertainty with respect to individuals' roles and students' feedback by means of Bayesian techniques; (ii) it allows us to calculate the optimal team configuration for the classroom in an enormous search space by using coalitional structure generation approaches that are supported by linear programming.

We tested the performance of the proposed tool in an experiment involving students that took part in three different team activities. The experiments suggest that the proposed tool is able to improve different teamwork aspects such as team dynamics and student satisfaction.

On the one hand, the results show that students perceived a higher degree of cooperation and coordination in teams proposed by the tool than those teams formed by classical methods. On the other hand, the experiment also provides greater support for students to perceive their teammates' attitudes as being positive. Despite these positive findings and the studies found in the research literature [53], we could not identify statistical differences with respect to decision-making tasks. We argue that this result may be influenced by the fact that the team activities were not oriented towards decision-making tasks and, therefore, differences in decision making may be more difficult to identify.

Finally, the experiments also suggest that the teams formed by our tool received more positive answers regarding student satisfaction aspects. In general, a higher percentage of students evaluated the team experience as positive, felt more comfortable working with teammates, perceived the work carried out in the team as satisfactory for all teammates, and evaluated their teammates higher.

We must note that although we used a random configuration for the initial team formation, we could include any questionnaire in order to have previous information about each student. In addition, the modularity of our proposal would allow us to use any other underlying grouping model such as

another role-based model in order to compare its effectiveness in educational contexts. We plan to deal with this issue in future works.

As for future lines of research, our current tool takes into account the behaviours of individuals on a team. However, there are other factors that we believe will have an impact on team performance. We consider interpersonal relationships to be a factor to be considered in future lines of work since they may affect team dynamics [100]. We also plan to include other types of factors such as student performance, level of knowledge, as well as constraints imposed by teachers.

Acknowledgements

This work is supported by the following projects: TIN2014-55206-R, TIN2012-36586-C03-01, PROMETEOII/2013/019, TIN2015-65515-C4-1-R, H2020-ICT-2015-688095.

References

- [1] E. Dunne, M. Rawlins, Bridging the gap between industry and higher education: Training academics to promote student teamwork, *Innovations in Education and Teaching International* 37 (4) (2000) 361–371.
- [2] S. G. Adams, B. Laksumanage, Building successful student teams in the engineering classroom, *Journal of STEM Education* 4 (3) (2003) 4.
- [3] P. M. Alexander, Virtual teamwork in very large undergraduate classes, *Computers & Education* 47 (2) (2006) 127–147.
- [4] T. A. Fredrick, Facilitating better teamwork: Analyzing the challenges and strategies of classroom-based collaboration, *Business Communication Quarterly*.
- [5] P. Sancho-Thomas, R. Fuentes-Fernández, B. Fernández-Manjón, Learning teamwork skills in university programming courses, *Computers & Education* 53 (2) (2009) 517–531.
- [6] G. Stahl, *Group cognition: Computer support for building collaborative knowledge*, Mit Press Cambridge, MA, 2006.

- [7] P. Tarricone, J. Luca, Employees, teamwork and social interdependence—a formula for successful business?, *Team Performance Management: An International Journal* 8 (3/4) (2002) 54–59.
- [8] J.-W. van Aalst, C. van der Mast, Performer: an instrument for multidisciplinary courseware teams to share knowledge and experiences, *Computers & Education* 41 (1) (2003) 39–48.
- [9] D. Fay, C. Borrill, Z. Amir, R. Haward, M. A. West, Getting the most out of multidisciplinary teams: A multi-sample study of team innovation in health care, *Journal of Occupational and Organizational Psychology* 79 (4) (2006) 553–567.
- [10] V. Ratcheva, Integrating diverse knowledge through boundary spanning processes—the case of multidisciplinary project teams, *International Journal of Project Management* 27 (3) (2009) 206–215.
- [11] N. L. Kerr, R. S. Tindale, Group performance and decision making, *Annu. Rev. Psychol.* 55 (2004) 623–655.
- [12] K. Behfar, R. Friedman, J. M. Brett, The team negotiation challenge: Defining and managing the internal challenges of negotiating teams, in: *IACM 21st Annual Conference Paper*, 2008.
- [13] C. K. De Dreu, L. R. Weingart, Task versus relationship conflict, team performance, and team member satisfaction: a meta-analysis., *Journal of applied Psychology* 88 (4) (2003) 741.
- [14] G. M. Parker, *Team players and teamwork: New strategies for developing successful collaboration*, John Wiley & Sons, 2011.
- [15] R. M. Belbin, *Team roles at work*, Routledge, 2010.
- [16] S. G. Fisher, W. Macrosson, G. Sharp, Further evidence concerning the Belbin team role self-perception inventory, *Personnel review* 25 (2) (1996) 61–67.
- [17] B. Senior, Team roles and team performance: is there ‘really’ a link?, *Journal of occupational and organizational psychology* 70 (3) (1997) 241–258.

- [18] D. Partington, H. Harris, Team role balance and team performance: an empirical study, *Journal of Management Development* 18 (8) (1999) 694–705.
- [19] A. Aritzeta, S. Swailes, B. Senior, Belbin's team role model: Development, validity and applications for team building*, *Journal of Management Studies* 44 (1) (2007) 96–118.
- [20] D. van Dierendonck, R. Groen, Belbin revisited: A multitrait-multimethod investigation of a team role instrument, *European Journal of Work and Organizational Psychology* 20 (3) (2011) 345–366.
- [21] J. Sommerville, S. Dalziel, Project teambuilding—the applicability of Belbin's team-role self-perception inventory, *International Journal of Project Management* 16 (3) (1998) 165–171.
- [22] D. Smith, Crisis management teams: issues in the management of operational crises, *Risk Management* (2000) 61–78.
- [23] M. Rajendran, Analysis of team effectiveness in software development teams working on hardware and software environments using Belbin self-perception inventory, *Journal of Management Development* 24 (8) (2005) 738–753.
- [24] E. C. King, The roles of student musicians in quartet rehearsals, *Psychology of music* 34 (2) (2006) 262–282.
- [25] R. J. Blignaut, I. M. Venter, Teamwork: Can it equip university science students with more than rigid subject knowledge?, *Computers & Education* 31 (3) (1998) 265–279.
- [26] S. M. Henry, K. Todd Stevens, Using belbin's leadership role to improve team effectiveness: an empirical investigation, *Journal of Systems and Software* 44 (3) (1999) 241–250.
- [27] H. van de Water, K. Ahaus, R. Rozier, Team roles, team balance and performance, *Journal of Management Development* 27 (5) (2008) 499–512.
- [28] A. Ounnas, H. C. Davis, D. E. Millard, A framework for semantic group formation in education, *Journal of Educational Technology & Society* 12 (4) (2009) 43–55.

- [29] V. Yannibelli, A. Amandi, A deterministic crowding evolutionary algorithm to form learning teams in a collaborative learning context, *Expert Systems with Applications* 39 (10) (2012) 8584–8592.
- [30] B. Senior, S. Swailes, A comparison of the belbin self perception inventory and observer's assessment sheet as measures of an individual's team roles, *International Journal of Selection and Assessment* 6 (1) (1998) 1–8.
- [31] T. Rahwan, T. P. Michalak, M. Wooldridge, N. R. Jennings, Coalition structure generation: A survey, *Artificial Intelligence* (2015) In Press.
- [32] R. M. Belbin, Team roles at work: A strategy for human resource management, zitiert in: *Teamarbeit und Teamentwicklung* (1993) 321.
- [33] Belbin, B. Associates, *The Belbin guide to succeeding at work*, A&C Black, 2009.
- [34] R. Meredith Belbin, *Management teams: Why they succeed or fail*, *Human Resource Management International Digest* 19 (3).
- [35] O. P. John, R. W. Robins, Accuracy and bias in self-perception: individual differences in self-enhancement and the role of narcissism., *Journal of personality and social psychology* 66 (1) (1994) 206.
- [36] D. L. Paulhus, O. P. John, Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives, *Journal of personality* 66 (6) (1998) 1025–1060.
- [37] E. Pronin, T. Gilovich, L. Ross, Objectivity in the eye of the beholder: divergent perceptions of bias in self versus others., *Psychological review* 111 (3) (2004) 781.
- [38] R. S. Hansen, Benefits and problems with student teams: Suggestions for improving team projects, *Journal of Education for Business* 82 (1) (2006) 11–19.
- [39] B. Senior, S. Swailes, A comparison of the Belbin self perception inventory and observer's assessment sheet as measures of an individual's team roles, *International Journal of Selection and Assessment* 6 (1) (1998) 1–8.

- [40] M. Higgs, U. Pewina, J. Ploch, Influence of team composition and task complexity on team performance, *Team Performance Management* 11 (7/8) (2005) 227–250.
URL <http://eprints.soton.ac.uk/51415/>
- [41] N. Ohta, V. Conitzer, R. Ichimura, Y. Sakurai, A. Iwasaki, M. Yokoo, Coalition structure generation utilizing compact characteristic function representations, in: *Principles and Practice of Constraint Programming - CP 2009*, Vol. 5732, 2009, pp. 623–638.
- [42] S. J. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach*, Prentice Hall, 2010.
- [43] K. F. Hirji, A comparison of exact, mid-p, and score tests for matched case-control studies, *Biometrics* (1991) 487–496.
- [44] A. Agresti, B. A. Coull, Approximate is better than “exact” for interval estimation of binomial proportions, *The American Statistician* 52 (2) (1998) 119–126.
- [45] R. G. Newcombe, Two-sided confidence intervals for the single proportion: comparison of seven methods, *Statistics in medicine* 17 (8) (1998) 857–872.
- [46] J. G. Hwang, M.-C. Yang, An optimality theory for mid p-values in 2×2 contingency tables, *Statistica Sinica* 11 (3) (2001) 807–826.
- [47] A. Agresti, M. Kateri, *Categorical data analysis*, Springer, 2011.
- [48] J. Graffelman, V. Moreno, The mid p-value in exact tests for hardy-weinberg equilibrium, *Statistical Applications in Genetics and Molecular Biology* 12 (4) (2013) 433–448.
- [49] N. P. Jewell, *Statistics for epidemiology*, CRC Press, 2003.
- [50] K. J. Rothman, S. Greenland, T. L. Lash, *Modern epidemiology*, Lippincott Williams & Wilkins, 2008.
- [51] J. E. Mathieu, T. S. Heffner, G. F. Goodwin, E. Salas, J. A. Cannon-Bowers, The influence of shared mental models on team process and performance., *Journal of applied psychology* 85 (2) (2000) 273.

- [52] E. Salas, N. J. Cooke, M. A. Rosen, On teams, teamwork, and team performance: Discoveries and developments, *Human Factors: The Journal of the Human Factors and Ergonomics Society* 50 (3) (2008) 540–547.
- [53] J. S. Prichard, N. A. Stanton, Testing Belbin’s team role theory of effective groups, *Journal of Management Development* 18 (8) (1999) 652–665.
- [54] J. A. Ross, The reliability, validity, and utility of self-assessment, *Practical Assessment, Research & Evaluation*.
- [55] E. D. Val, J. M. Alberola, V. Sánchez-Anguix, A. Palomares, M. D. Teruel, A Team Formation Tool for Educational Environments, in: *Trends in Practical Applications of Heterogeneous Multi-Agent Systems*, Vol. 293, 2014, pp. 173–181.
- [56] K. A. Bantel, S. E. Jackson, Top management and innovations in banking: does the composition of the top team make a difference?, *Strategic Management Journal* 10 (1989) 107.
- [57] S. E. Jackson, Team composition in organizational settings: Issues in managing an increasingly diverse work force., in: *Symposium on Group Productivity and Process, 1989*, Texas A & MU, College Station, TX, US, Sage Publications, Inc, 1991.
- [58] C. K. De Dreu, L. R. Weingart, Task versus relationship conflict, team performance, and team member satisfaction: a meta-analysis., *Journal of applied Psychology* 88 (4) (2003) 741.
- [59] M. D. Ensley, K. M. Hmieleski, A comparative study of new venture top management team composition, dynamics and performance between university-based and independent start-ups, *Research Policy* 34 (7) (2005) 1091–1105.
- [60] J. Mathieu, M. T. Maynard, T. Rapp, L. Gilson, Team effectiveness 1997-2007: A review of recent advancements and a glimpse into the future, *Journal of Management* 34 (3) (2008) 410–476.
- [61] M. André, M. G. Baldoquín, S. T. Acuña, Formal model for assigning human resources to teams in software projects, *Information and Software Technology* 53 (3) (2011) 259–275.

- [62] H. Spoelstra, P. van Rosmalen, T. Houtmans, P. Sloep, Team formation instruments to enhance learner interactions in open learning environments, *Computers in Human Behavior* 45 (2015) 11 – 20.
- [63] D. Francescato, M. Mebane, R. Porcelli, C. Attanasio, M. Pulino, Developing professional skills and social capital through computer supported collaborative learning in university contexts, *International Journal of Human-Computer Studies* 65 (2) (2007) 140–152.
- [64] S. Dewiyanti, S. Brand-Gruwel, W. Jochems, N. J. Broers, Students' experiences with collaborative learning in asynchronous computer-supported collaborative learning environments, *Computers in Human Behavior* 23 (1) (2007) 496–514.
- [65] A. K. Tolmie, K. J. Topping, D. Christie, C. Donaldson, C. Howe, E. Jessiman, K. Livingston, A. Thurston, Social effects of collaborative learning in primary schools, *Learning and Instruction* 20 (3) (2010) 177–191.
- [66] B. Schneider, P. Jermann, G. Zufferey, P. Dillenbourg, Benefits of a tangible interface for collaborative learning and interaction, *Learning Technologies, IEEE Transactions on* 4 (3) (2011) 222–232.
- [67] L. Blasco-Arcas, I. Buil, B. Hernández-Ortega, F. J. Sese, Using clickers in class. the role of interactivity, active collaborative learning and engagement in learning performance, *Computers & Education* 62 (2013) 102–110.
- [68] S. Graf, R. Bekele, Forming heterogeneous groups for intelligent collaborative learning systems with ant colony optimization, in: *Proceedings of the 8th international conference on Intelligent Tutoring Systems*, Springer, 2006, pp. 217–226.
- [69] D.-Y. Wang, S. S. Lin, C.-T. Sun, Diana: A computer-supported heterogeneous grouping system for teachers to conduct successful small learning groups, *Computers in Human Behavior* 23 (4) (2007) 1997 – 2010.
- [70] E. Grigorenko, R. Sternberg, Styles of thinking, abilities, and academic performance, *Exceptional children* 63 (3) (1997) 295–312.

- [71] P. Bergey, M. King, Team machine: A decision support system for team formation, *Decision Sciences Journal of Innovative Education* 12 (2) (2014) 109–130.
- [72] Y.-T. Lin, Y.-M. Huang, S.-C. Cheng, An automatic group composition system for composing collaborative learning groups using enhanced particle swarm optimization, *Computers & Education* 55 (4) (2010) 1483 – 1493.
- [73] R. Cavanaugh, M. Ellis, R. Layton, M. Ardis, Automating the process of assigning students to cooperative-learning teams, in: *Proceedings of the 2004 ASEE Annual Conference*, 2004.
- [74] J. Moreno, D. A. Ovalle, R. M. Vicari, A genetic algorithm approach for group formation in collaborative learning considering multiple student characteristics, *Computers & Education* 58 (1) (2012) 560–569.
- [75] C. E. Christodouloupoulos, K. A. Papanikolaou, A group formation tool in an e-learning context, *2012 IEEE 24th International Conference on Tools with Artificial Intelligence* 2 (2007) 117–123.
- [76] H. Wi, S. Oh, J. Mun, M. Jung, A team formation model based on knowledge and collaboration, *Expert Systems with Applications* 36 (5) (2009) 9121 – 9134.
- [77] G. L. Stewart, I. S. Fulmer, M. R. Barrick, An exploration of member roles as a multilevel linking mechanism for individual traits and team outcomes, *Personnel Psychology* 58 (2) (2005) 343–365.
- [78] J. E. Mathieu, S. I. Tannenbaum, M. R. Kukenberger, J. S. Donsbach, G. M. Alliger, Team role experience and orientation a measure and tests of construct validity, *Group & Organization Management* 40 (1) (2015) 6–34.
- [79] D. G. Ancona, D. F. Caldwell, Beyond task and maintenance defining external functions in groups, *Group & Organization Management* 13 (4) (1988) 468–494.
- [80] E. Sundstrom, K. P. De Meuse, D. Futrell, Work teams: Applications and effectiveness., *American psychologist* 45 (2) (1990) 120.

- [81] J. Fransen, P. A. Kirschner, G. Erkens, Mediating team effectiveness in the context of collaborative learning: The importance of team and task awareness, *Computers in Human Behavior* 27 (3) (2011) 1103–1113.
- [82] R. F. Bales, *Interaction process analysis; a method for the study of small groups*.
- [83] C. Margerison, D. McCann, *Team management systems: How to lead a winning team* (1985).
- [84] G. M. Parker, *Cross-functional teams: Working with allies, enemies, and other strangers*, John Wiley & Sons, 2003.
- [85] S. E. Humphrey, F. P. Morgeson, M. J. Mannor, Developing a theory of the strategic core of teams: a role composition model of team performance., *Journal of Applied Psychology* 94 (1) (2009) 48.
- [86] J. Bais, J. Giletto, J. Balmaceda, S. Schiaffino, J. A. Díaz, An assistant for group formation in cscl based on constraint satisfaction, *Proceedings of ASAI* (2013) 12–23.
- [87] J. M. Balmaceda, S. N. Schiaffino, J. A. D. Pace, Using constraint satisfaction to aid group formation in cscl., *Revista Iberoamericana de Inteligencia Artificial* 17 (53) (2014) 35–45.
- [88] F. Mumma, *Team-work & team-roles*. mumma f, ed. <http://teambuilding.com>, <http://hrdq.com>.
- [89] W. Park, H. Bang, Team role balance and team performance, in: *Belbin Biennial Conference, “Changing Role of Management in the 21st Century,”* Clare College, Cambridge, England, 2002.
- [90] P. Jefferies, F. Grodzinsky, J. Griffin, Building successful on-line learning communities across international boundaries: a case study, *ETHICOMP*, 2004.
- [91] S. McDonald, H. M. Edwards, Who should test whom?, *Communications of the ACM* 50 (1) (2007) 66–71.
- [92] A. Furnham, H. Steele, D. Pendleton, A psychometric assessment of the Belbin team-role self-perception inventory, *Journal of Occupational and Organizational Psychology* 66 (3) (1993) 245–257.

- [93] W. G. Broucek, G. Randell, An assessment of the construct validity of the Belbin self-perception inventory and observer's assessment from the perspective of the five-factor model, *Journal of Occupational and Organizational Psychology* 69 (4) (1996) 389–405.
- [94] V. Yannibelli, A. Amandi, A memetic algorithm for collaborative learning team formation in the context of software engineering courses, in: *Advances in New Technologies, Interactive Interfaces and Communicability*, Springer, 2012, pp. 92–103.
- [95] R. Fares, R. Costaguta, A multi-agent model that promotes team-role balance in computer supported collaborative learning, in: *Advances in New Technologies, Interactive Interfaces and Communicability*, Springer, 2012, pp. 85–91.
- [96] J. M. Alberola, E. del Val, V. Sanchez-Anguix, V. Julian, Simulating a collective intelligence approach to student team formation, in: *Hybrid Artificial Intelligent Systems*, Springer, 2013, pp. 161–170.
- [97] H.-Y. Ku, H. W. Tseng, C. Akarasriworn, Collaboration factors, teamwork satisfaction, and student attitudes toward online collaborative learning, *Computers in Human Behavior* 29 (3) (2013) 922–929.
- [98] E. Pfaff, P. Huddleston, Does it matter if I hate teamwork? what impacts student attitudes toward teamwork, *Journal of Marketing Education* 25 (1) (2003) 37–45.
- [99] G. R. Jones, J. M. George, The experience and evolution of trust: Implications for cooperation and teamwork, *Academy of management review* 23 (3) (1998) 531–546.
- [100] I. B. Myers, M. H. McCaulley, R. Most, *Manual, a guide to the development and use of the Myers-Briggs type indicator*, Consulting Psychologists Press, 1985.