Comparing computational algorithms for team formation in the classroom: a classroom experience

Victor Sanchez-Anguix[1] · Juan M. Alberola[2] · Elena Del Val[2] · Alberto Palomares[2] · Maria Dolores Teruel[3]

Abstract

Throughout recent years, several researchers have proposed computational tools and algorithms to support team formation in the classroom. The result is that team formation algorithms have been widely applied in classroom environments to create well-balanced teams. One of the challenges in designing algorithms for automatic team formation is designing an appropriate function to estimate team performance, which is used as part of the optimization algorithm that divides students into teams. This function (referred to as a team evaluation heuristic) serves as an approximation to team performance, which is a complex phenomenon that is difficult to quantitatively assess in many settings and that cannot be accurately calculated prior to the task at hand. Despite showing their relative success compared to traditional and manual team formation strategies (manually employed by lecturers and teachers), there is a lack of research comparing team evaluation heuristics in a real classroom setting. Such a comparison would help teachers, practitioners, and system designers to appropriately select the most suitable team formation algorithms. In this article, we present an experimental evaluation that was carried out in a Bachelor's Degree Program in Tourism that compares two team evaluation heuristics based on Belbin and Myer-Briggs. The experimental evaluation was carried out by means of an intelligent, extensible team formation tool whose optimization is based on an integer linear model that can be extended to support different team evaluation heuristics.

Keywords Team formation · Artificial intelligence · Education · Decision support systems · Team assembly

1 Introduction

Teamwork related skills have emerged as one of the most prominent competencies to be developed in Higher Education [2, 11, 12, 16, 49, 52]. As a consequence, many courses have included team activities in their curricula to hone this general competency [30, 39, 46]. This trend responds to the widely accepted reality that teams are a major component in a vast majority of modern companies [14, 38, 45, 54, 57]. The reality is that, despite the difficulty of working as a team [8, 29], teams have proven to be successful when dealing with the increasingly complex problems that are faced in the current work environment. Hence, the importance of acquiring teamwork skills and attitudes.

In the classroom, team activities also have an underlying learning objective that needs to be accomplished. As highlighted by some studies, appropriate team dynamics should emerge to reinforce teamwork and to foster learning among students [51, 59]. If not, students may focus on the negative consequences and feelings that emerge from negative team dynamics, deterring them from focusing on the learning task and honing teamwork skills. Therefore, it is important to provide an appropriate team environment. Team formation strategies may help to implement such an environment.

The problem of computationally assisting in forming a team is not new and has received the attention of different computer science disciplines. These have provided different system solutions to the problem. In the literature, these problems have been regarded as either team assembly or team formation problems [22, 28]. In this article, we use both terms interchangeably. Recently, [22] classified systems supporting team assembly based on users' agency and users' participation, resulting in four types of systems: *optimized teams*, *staffed teams*, *self-assembled teams*, and *augmented teams*. This article focuses on *optimized teams*, i.e., systems where teams are automatically formed based on algorithms that follow a varied plethora of criteria.

✉ Victor Sanchez-Anguix
   vsanchez@eio.upv.es

Extended author information available on the last page of the article

Systems that support the team assembly problem by providing algorithms for the automatic formation of optimized teams have traditionally been studied from the perspective of artificial intelligence and optimization [28]. Team formation problems have two associated challenges. On the one hand, many team formation problems, even in their simplest variant, are considered NP problems [28]. Therefore, one of the challenges is to develop efficient algorithms (exact or approximate), that can find the optimal set of teams based on some objective function. On the other hand, the other challenge is to design an appropriate function that estimates the performance of a set of teams in the real world. In many settings, this challenge boils down to the design of a function that estimates the real performance of a team for a specific task. As the reader may suspect, this is not a trivial task by any means. Team performance is difficult to assess and define, as it is affected by many (sometimes qualitative) factors like results, dynamics, work environment, long-term perspectives, synergies, etc. As a result, this function will always be an estimation of team performance in the real world, and many authors propose proxy functions for approximating team performance that are inspired by organizational theory, psychology, educational theories, and so forth. We refer to these functions as team evaluation heuristics since they act as proxy functions for the performance of a team in a real setting.

In the context of education, one of the most prominent factors used as a team evaluation heuristic is personality and behavior [21, 35, 41, 56]. This is explainable since personality has been shown to influence cooperation [33]. From the different taxonomies and studies that focus on describing personality, many authors have focused on proposing team evaluation heuristics that are either inspired by Belbin's role taxonomy [9] or by the Myers-Briggs type indicator [40] (MBTI). The Belbin taxonomy aims to identify individual behavioral patterns that arise in successful teams. The Myers-Briggs is a popular assessment scale for personality types across four dimensions. In fact, the two aforementioned theories, Belbin's theory and Myer-Briggs theory, are perhaps the two most widespread theories that have influenced the design of computer-enabled tools and algorithms for team formation.

Several proposals have studied the performance of team evaluation heuristics inspired by Belbin or MBTI when compared to traditional team formation strategies that are manually carried out by lecturers. These include random teams, teams that are self-assembled by students, and teams that are manually created based on the instructors' expertise [1, 6, 7, 31, 50, 62]. Despite this, there is still a lack of research comparing the team evaluation heuristics proposed in the literature for optimized team formation. In other words, even though there is individual support for the use of some of these team evaluation heuristics versus traditional

criteria, there is no specific research on which team evaluation heuristics most positively influence team dynamics in the classroom. Therefore, there is a need for studies that compare the team evaluation heuristics inspired by Belbin's theory and Myer-Briggs theory, These studies may help to determine similarities and differences in team performance and learning.

In addition, many of the studies found in the literature focus on the effectiveness of the optimization algorithms designed for team formation in the classroom, or they focus on showing the benefits of a team evaluation heuristic versus traditional team formation strategies such as random assignment or assignment that is based on the teachers' expertise. Nevertheless, they offer little insight into how these computer-enabled tools should be applied in practice, the problems that may arise when using these tools, the strategies that work best for managing this type of experience, and what the main advantages and disadvantages are of using these types of tools.

This article aims to help resolve some of these issues. The **main contribution** of this article is that we design and compare two team evaluation heuristics that are based on two of the most popular theories in the optimized team formation literature in a real classroom experiment. One is based on Belbin's role taxonomy and the other is based on the Myers' Briggs type indicator. As mentioned, this is one of the gaps in the state of the art on team formation problems in the classroom. Both team evaluation heuristics aim for heterogeneity on the team, be it roles or personality traits. The reasons why we focus on comparing two team evaluation heuristics inspired by Belbin's role taxonomy and Myer-Briggs theory are that they are two of the most widespread theories when designing algorithms for team formation and different authors have provided individual evidence that strategies based on these two theories may provide better team dynamics than teams formed manually by lecturers based on their own traditional criteria [1, 3, 6, 31, 36]. The evaluation was conducted in the Bachelor's Degree Program in Tourism at Universitat Politècnica de València over five academic years (2014-2019). To the best of the authors' knowledge, this is the first study that explicitly compares two team evaluation heuristics inspired by Belbin and Myers-Briggs despite their popularity in optimized team formation approaches. In practical terms, studies that compare different team evaluation heuristics provide insights and guidelines that lecturers may use to form teams in their classrooms. In addition, this article **also contributes to the state of the art** by presenting an intelligent team formation tool based on an integer linear programming model that allows us to extend it to support a wide range of team evaluation heuristics. The tool provides capabilities for lecturers to analyze the performance of their team activities. Finally, we include interviews with lecturers with insights about the experience and the benefits and

problems that may arise when using this type of tool. This will contribute to understanding the practical issues that arise when using these tools in the classroom.

The remainder of this paper is organized as follows. In Section 2, we provide a thoughtful review of the models and tools used for automatized team formation in the classroom. Section 3 presents the different components of the team formation tool that was used for the experiments, including its design, the mathematical formulation of the problem, and the proposal of the two team evaluation heuristics. Section 4 describes the experiments that we carried out and the most relevant findings from the results obtained. Finally, in Section 5, we provide some concluding remarks and plans for future work.

2 Related work

In this section, we mainly focus on analyzing contributions that have used a team evaluation heuristic inspired by either Belbin or by Myer-Briggs. For more detailed and broad reviews on the topic of algorithmic team formation in the classroom, the readers are directed to the following reviews [10, 20, 32, 35, 41].

Alberola et al. [1] designed an artificial intelligence algorithm based on integer programming and Bayesian learning that creates teams by grouping students with heterogeneous Belbin roles. The strategy was found to perform better in many areas than randomly grouping students into teams, showing how Belbin's theory may be beneficial in forming teams in the classroom. The algorithms assume that several team activities are carried out so that feedback can be gathered on the role played by each individual on the teams. Similar to their proposal, our proposal also uses integer programming as an optimization technique. However, the heuristics implemented in our tool focus on one-shot team activities. In other words, they do not require previous team activities to gather information and data about the participants. [62–65] have proposed several metaheuristics to form well-balanced teams in the classroom. More specifically, the authors propose a team formation heuristic that penalizes teams that lack some Belbin roles or teams where roles are over-represented. In [65], the performance of a hybrid evolutionary algorithm is tested and compared with previous algorithm versions based on memetic and genetic algorithms. However, to the best of our knowledge, the experiments focus on comparing the performance of the optimization algorithms and not on assessing the quality of the team formation heuristic in a classroom. In another work, [17] propose a team formation methodology that is based on sociograms that detect relationships between students and the teacher's own judgment on Belbin's taxonomy. The proposed tool detects triads on sociograms. Then, triads are provided to lecturers,

who can change teams according to their expertise and Belbin's theory. The methodology was only tested in a small pilot experiment involving a limited number of students, and it was not compared with other team formation criteria. [31] propose a tool that generates the most promising team structure for a classroom. It takes into account the demographics of the students and the number of Belbin roles present on the team, and it attempts to ensure homogeneity among the learning styles of the students. The proposed strategy was tested with success against self-formed teams, manually formed teams, and a simplified version of the grouping criteria that is only based on learning styles. However, the team formation heuristic is not compared with other team formation algorithms based on other principles such as MBTI. [42] propose a framework for team formation which is based on semantic web technologies and constraint satisfaction optimization. This framework is aimed at providing support for lecturers when forming teams based on Belbin roles and other constraints, such as the number of females per group. The authors evaluate this proposal by running different simulated classes of students, but no comparison is carried out in a real classroom with other team formation algorithms. Recently, [7] and [58] have proposed a genetic algorithm named *TalSoR*, whose fitness function fosters teams that are balanced with respect to Belbin's role taxonomy and gender representation. The proposed genetic algorithm has been used in a real course in Chemical Engineering for two academic years, where students engage in project-based learning. The proposed team evaluation heuristic is compared with the results obtained by self-assembled teams from previous academic years. The results show that the proposed heuristic and the optimization algorithm helped the students obtain higher marks in the course, increased their interest in the topic by the end of the course, and provided a more positive perception of the team experience. Finally, [18] carried out a classroom experience to examine whether or not organizing balanced teams according to Belbin's role taxonomy was linked to students' performance in a business fundamentals course. To do this, they carried out a posterior analysis of teams that were self-assembled by the students and compared the performance of balanced teams with respect to the performance of unbalanced teams. The results show that, in general, the students' performance tends to be less extreme on balanced teams than on unbalanced teams.

Chen and Lin [13] present a team formation algorithm based on the compatibility relationships among MBTI personalities, the teamwork capabilities, and the knowledge of team members. Different from other works, the goal in the aforementioned article is not to divide a classroom into teams but rather to form the most effective team for an engineering project. Thus, only a single team is needed and provided by the proposed method. A case study is also presented with synthetic data, but no real evaluation of the algorithm is car-

ried out. Another study that is based on MBTI is presented in [15]. In this case, the authors aim to study the impact of the personality type of team members in the formation of self-assembled software engineering teams. The MBTI types of contributors to the Python Enhancement Proposal are estimated by means of text analysis on social networks. Then, the authors carried out simulation experiments to understand what characteristics helped to predict team composition, with some MBTI dimensions found to be relevant in predicting team composition. This work differs from other reviewed works since its aim is not explicitly forming optimal teams in a classroom but rather understanding some of the factors that may help the formation of self-assembled teams. Andrejczuk et al. [6] and [4] propose an algorithm that is based on finding teams that are proficient for a given task and whose congeniality is also high. In this work, congeniality is based both on personality diversity and gender diversity. In the case of personality diversity, the team evaluation heuristic is inspired by a reduced version of MBTI. The authors propose two team formation algorithms, one based on an integer linear programming (ILP) formulation that is capable of finding the optimal solution and an anytime heuristic based on local search. The team evaluation heuristic was tested using several types of tasks in the classroom comparing its results with those of the traditional criteria used by lecturers in the classroom. The results suggest that the teams formed by their algorithm are better performing than those created by traditional lecturers' strategies. However, no comparison is carried out with other team evaluation heuristics. Varvel et al. [60] use MBTI for team formation in different courses such as Computer Engineering, Construction Management, or Chemical Engineering. This study does not conclude that a particular combination of personality types has a more positive impact on team performance as other studies do. However, knowing these types helps individuals to improve their communication skills, levels of trust, and other characteristics that influence team performance. Amato and Amato [3] analyze the relationship between team learning perceptions and communication styles by using MBTI. Specifically, this study compares students' satisfaction when working with people with similar personality types with people who have complementary personality types. This study does not provide any general conclusion since they found different preferences depending on the course and the subject. Mazni et al. [36] use MBTI to form teams and analyze their performance in software engineering courses. This study concludes that the combination of personality types is key in determining the performance of the team. According to the authors, heterogeneous teams were more successful when developing high-quality software. In contrast, homogeneous teams were more suitable for less challenging projects.

Other authors have focused on the task of forming teams in online settings, where normal team dynamics may be affected by physical distance and lack of familiarity with team members. The work presented in [50] presents a clustering approach to team formation that takes into consideration the activity of online learners on a MOOC platform. The clustering algorithm aims to group together students that have similar activity on the platform, with the goal of creating teams with similar learning aspirations. The clustering algorithm was compared with an algorithm that assigns teams randomly. The results show how the clustering algorithm promoted active teams in online settings as well as student satisfaction. In [53], the authors propose a set of three rule-based team formation algorithms that aim to create productive, creative, or learning teams in MOOC environments. The participants were involved in a questionnaire to provide feedback on the hypothesis used by each of the three rule-based algorithms mentioned above to create different alternatives of teams, with the participants validating the hypothesis behind the creation of productive and learning teams. In addition to this, the participants validated the logic behind the team formation mechanisms. However, no experiments were carried out to assess the real performance of teams formed by the proposed algorithms. Another interesting study is presented in [19], where the authors present a multi-objective genetic algorithm based on NSGA-II to form teams that maximize the heterogeneity within teams and maximize the homogeneity among teams. The attributes used to group students into teams include gender, an indicator of leadership, and the communication skills of the students. However, it should be highlighted that these characteristics were synthetically created and they do not include information about real students nor was the quality of the grouping criteria tested in a real classroom. To a lesser extent, other authors have used team evaluation heuristics inspired by other personality instruments like Big Five [26]. For instance, [48] propose a genetic algorithm for grouping students based on a team evaluation heuristic inspired by the Big Five inventory. The genetic algorithm aims to find balanced teams based on their personality types and the personality types present in the classroom. After optimizing the hyperparameters of the genetic algorithm, the authors carried out a real experiment in the classroom to assess whether or not teams formed by the genetic algorithm outperformed teams assembled by the students. The experiments show that higher grades were achieved by teams formed with the team evaluation heuristic proposed by the authors. The authors in [34] propose an integer linear programming model for forming teams centered around a team leaders. First of all, the authors propose a method that considers multiple characteristics, including the Big Five personality traits, to assess the value of a student as a leader. Once the best leaders are identified, the authors use an integer linear programming model to group students around each leader. To do this, the authors consider several characteristics and the congeniality of each team member with

the assigned leader to define their team evaluation heuristic. Different configurations of the model are evaluated against self-assembled teams. The results suggest that groups formed via this model are less prone to extreme cases (i.e., they show less deviation).

A brief overview of the analysis carried out in this section can be found in Table 1. First, as the reader may observe, inspiring team evaluation heuristics on both Belbin's taxonomy and MBTI is a consistent trend since the first works on team formation problems. Second, while many team evaluation heuristics have been used with different degrees of success compared to traditional grouping criteria (i.e., random teams, self-assembled teams, teams formed by the instructor), there is a lack of research comparing the team evaluation heuristics proposed in the literature in a real classroom setting. More specifically, although there is individual support for using team evaluation heuristics based on Belbin's role taxonomy or MBTI personality types [1, 3, 6, 7, 18, 31, 36] compared to traditional grouping criteria, there is a lack of research comparing the two families of team evaluation heuristics. This is partly due to the complexity of setting an experiment involving a large number of teams as well as the large number of teams and activities that are necessary to compare several criteria with statistical significance. Nevertheless, we argue that it is important to compare the team performance achieved by different team evaluation heuristics since the goal is to obtain the best performing teams and to provide insights that lecturers can use to form the best teams in their classrooms. This is clearly a gap that needs to be addressed. The main contribution of this article helps to fill this gap by comparing two team evaluation heuristics inspired by two of the most common criteria, Belbin's role taxonomy and Myers-Briggs type indicator, in a real class environment.

3 Team formation tool

In this section, we describe the design and implementation of the proposed tool for team formation in classroom environments. This tool uses optimization algorithms for team formation and is based on a previous development presented in [1].

In the current version, the automatic team formation is supported by two components: an optimization model, and two heuristics that drive the optimization model and that allow making either a Belbin-based team formation or an MBTI-based team formation. First, we present a general description of the capabilities as well as the general architecture of the tool. Afterwards, we describe the optimization model based on integer programming that allows us to divide a classroom into teams based on a heuristic to evaluate team

Table 1 Analysis of team formation algorithms based on the application context, the grouping criteria, the type of evaluation, and if the obtained team performance has been compared against traditional grouping criteria manually carried out by lecturers or other team formation algorithms

	Context	Team evaluation heuristic	Evaluation	Comparison vs. traditional criteria	Comparison with other team evaluation heuristics
[13]	Experimental	Belbin	Simulation		
[60]	Classroom	MBTI	Classroom		
[3]	Classroom	MBTI	Classroom	x	
[42]	Classroom	Belbin	Simulation		
[36]	Classroom	MBTI	Classroom	x	
[31]	Classroom	Belbin and learning styles	Classroom	x	x
[53]	MOOC	–	Classroom	x	
[1]	Classroom	Belbin	Classroom	x	
[15]	Classroom	MBTI	Simulation		
[50]	MOOC	–	Classroom	x	
[17]	Classroom	Belbin	Classroom		
[62–65]	Classroom	Belbin	Simulation		
[4, 6]	Classroom	MBTI	Classroom	x	
[19]	Experimental	–	Simulation		
[48]	Classroom	Big Five	Classroom	x	
[18]	Classroom	Belbin	Classroom	x	
[34]	Classroom	Big Five	Classroom	x	
[7, 58]	Classroom	Belbin	Classroom	x	

quality. Finally, we describe the two team evaluation heuristics that are implemented in the tool and that can be plugged into the optimization model to obtain different team formations.

3.1 Tool capabilities and architecture

The team formation tool is divided in a front-end for users' interaction and a back-end for team formation calculation and data visualisation (Fig. 1).

The front-end is focused on users' interaction. Specifically, the tool provides the following functionalities for lecturers: (i) registration and login; (ii) register a classroom for a team activity; (iii) create a team activity; (iv) activate predefined questionnaires or create new ones before a team activity; (v) run the team formation service; (vi) activate or create post-activity questionnaires; (vii) visualise the information

For students, the tool provides the following functionalities: (i) registration and login; (ii) complete the questionnaires provided before a team activity; (iii) share information with teammates during a team activity; (iv) complete the post-activity questionnaire; (v) visualise the results of the questionnaires.

Some of the functionalities mentioned above are integrated in the Learning Management System (LMS) of the university due to privacy concerns and because the LMS already provides several facilities (e.g. user authentication, authorisation, or questionnaire management). Since we used two different heuristics, we incorporated the Belbin Self-Perception Inventory and the MBTI as predefined questionnaires that should be completed before the team activity. In addition, we also provided a post-activity questionnaire focused on measuring the students' satisfaction as well as the team dynamics.

The back-end is composed by the team formation service, the data visualisation service, and the database. First, by taking into account the information gathered by the questionnaires as well as other parameters (such as the number of students per team), the team formation service is ready to generate the proposal of teams. The implementation used in this paper incorporates two different heuristics (i.e., the one based on Belbin's roles and the MBTI). However, as mentioned, the extensibility of this tool easily allows the integration of other heuristics in the future as described in Section 3.2. Second, the data visualisation service allows users to visualise some data, such as the most predominant Belbin roles, the MBTI personalities, or the satisfaction of each team. The back-end runs on our own servers. Finally, the database stores all of the information gathered, such as the teams, the questionnaires, or the answers of each student.

3.2 Optimization model

As mentioned previously, automatic team formation is supported by a general optimization model and two team evaluation heuristics, which can be easily replaced by others. The optimization model is based on Integer Linear Programming (ILP) [61] implemented using CPLEX, although we plan to include support for other free solvers that facilitate the dissemination and use of the tool. The optimization model has been designed to make it extensible. Specifically, as the reader will observe, new team evaluation heuristics can be introduced in the ILP model by just substituting the function that accompanies each variable in the objective function. Therefore, the same ILP model can be used for testing a wide variety of criteria transparently. In our current implementation, we have implemented two team evaluation heuristics: one based on Belbin, and another one based on MBTI.

From this point on, we provide the mathematical programming formulation for the team formation problem in the classroom, based on ILP. Let $\mathcal{S} = \{s_1, s_2, \ldots, s_n\}$ represent the set of students in a classroom with n individuals. We assume that not all of the team sizes are appropriate for the activity to be carried out by the students. Then, we define a minimum and a maximum team size: l and u.

We also assume that the lecturers may have specific knowledge about the class dynamics. This knowledge can be

Fig. 1 Overall architecture of the team formation tool
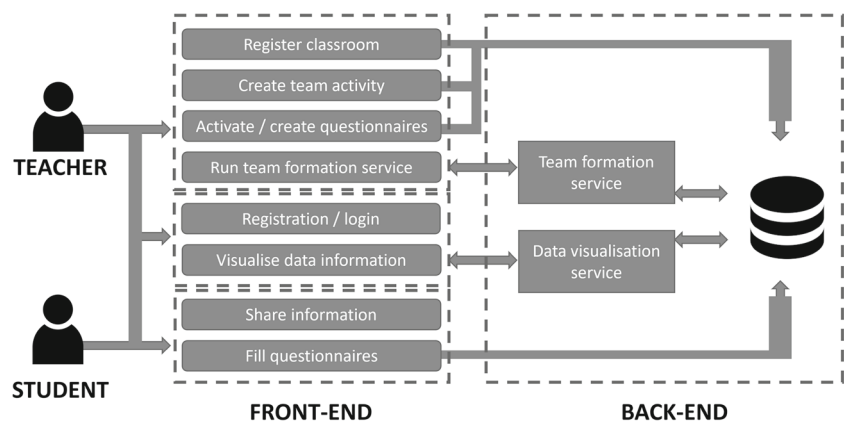
translated into rules about what teams should be avoided. Specifically, we say that a pair of students (s_i, s_j) is incompatible when the students cannot be part of the same team due to academic or personal reasons (e.g., conflict, different languages, etc.). Additionally, we define $\mathcal{N}$ as the set that contains all of the pairs of incompatible students. Based on this definition and the minimum and maximum team size, one can provide additional definitions. Let $t_i \subset \mathcal{S}$ represent a team of students, and $|t_i|$ represent the size of that team. We use the notation $s_j \in t_i$ to denote that the student s_j participates in team t_i. We say that a team t_i is feasible when $l \leq |t_i| \leq u$, and $\forall s_j, s_k \in t_i, (s_j, s_k) \notin \mathcal{N}$. From this point on, we will only focus on feasible teams, and we represent the set of all feasible teams as $\mathcal{T}$. Similarly, the lecturers may also desire to put certain pairs of students together on a team. We say that a pair of students (s_i, s_j) is compulsory when the students must be part of the same team. We define $\mathcal{C}$ as the set of compulsory pairs of students that must be present on selected teams. Later on, we describe the family of constraints that ensure that all of the compulsory pairs are present on the teams proposed.

The problem of forming teams in a classroom consists of deciding the disjoint teams in which a classroom is partitioned. For each feasible team t_i, we decide whether or not that particular team is to be used in the classroom. Therefore, for each team t_i we define a binary decision variable δ_i that represents whether or not the team t_i is used in the classroom (1 if it is used, 0 otherwise). Given the definition of these decision variables, we formalize the team formation problem in the classroom as follows:

$$\max Z = \sum_{t_i \in \mathcal{T}} f(t_i) \times \delta_i$$

subject to:

[student j in a team]:
$$\sum_{t_i \in \mathcal{T}, s_j \in t_i} \delta_i = 1; \ j = 1, \ldots, n$$

[compulsory pair j and k]:
$$\sum_{t_i \in \mathcal{T}; s_j, s_k \in t_i} \delta_i = 1; \ \forall (s_j, s_k) \in \mathcal{C}$$

where $f : \mathcal{T} \rightarrow \mathbf{R}$ is the team evaluation heuristic, which serves as a proxy function for team performance. As the reader can observe, this function f is general, and it can represent any function that, given a team and the information related to its members, obtains a real number. Thus, this function can represent team evaluation heuristics based on Belbin's theory, Myer-Briggs, students' marks, learning styles, gender balance, or any other criteria (even a combination of those mentioned). The model aims to maximize the sum of the idoneity of each of the selected teams. As the reader may observe, the formulation is generic and can be adapted to any team formation criteria and function that estimates team idoneity. It should be noted that the formulation of the problem is linear, since $f(.)$ calculates a scalar value that is computed for each feasible team before the optimization problem is solved and it only depends on the team itself.

Now, let us describe the constraints of the optimization model. On the one hand, there is a constraint for each student in the classroom, represented by the family of constraints *student j in a team*. Each of these constraints ensures that each student takes part in exactly one of the selected teams. On the other hand, we have another family of constraints, *compulsory pair j and k*. There is exactly one constraint for each compulsory pair of students. The constraint ensures the selection of exactly one team where students s_j and s_k are together, thus forcing the membership of the pair on the same team.

There are two main reasons why we decided to focus our current tool on an ILP model. First, the ILP model proposed in this section is general and is independent of the team evaluation heuristic on which it is based. This allows us to easily reuse the same optimization algorithm for different team evaluation heuristics. Second, unlike other common optimization algorithms used in team formation like genetic algorithms or local search metaheuristics, ILP models guarantee obtaining the optimal solution according to the specified team evaluation heuristics while approximate algorithms do not. Our main goal in this article is to compare two team evaluation heuristics. Hence, we need to make sure that observed differences are due to differences in the quality of the team evaluation heuristics and not due to differences introduced by the stochastic behavior and approximation of heuristics and metaheuristics. Thus, despite being less scalable than approximate algorithms, we strongly recommended using an ILP model when comparing team evaluation heuristics.

Compared to other state-of-the-art ILP models for team formation in the classroom, there are some differences. For instance, the ILP model proposed by [34] does not evaluate the performance of the whole team but rather the congeniality of each team member with its leader, leaving aside the interactions among team members. Our model also allows forcing the inclusion of pairs of students on the same team based on to the instructor's criteria. This is also another difference when compared to [4]. In addition, we can also forbid pairs of students, and we can allow several and varied team sizes.

3.3 Team evaluation heuristics

As mentioned in Section 3.2, the linear model aims to maximize the performance of the teams formed in the classroom, i.e. $\sum_{t_i \in \mathcal{T}} f(t_i) \times \delta_i$. Of course, it is not feasible to assess the performance of a team exactly before its formation and action. Therefore, scholars have proposed different heuristics to roughly approximate the performance of a team based on a variety of criteria. As we mentioned, some of these criteria are based on psychological theories (e.g., MBTI personalities), management theories (e.g., Belbin roles), or academic experience (e.g., the distribution of skills in teams). Prospective heuristics are varied, and they may also take into consideration other factors such as the type of task at hand, previous project work, and even other psychological traits. The function $f(.)$ used in the optimization model represents any heuristic that estimates the performance or quality of a team based on any criteria.

In this work, we have implemented and compared two heuristics: one based on MBTI, and another one based on Belbin roles. Nevertheless, the reader should be aware that the optimization model is general and can be adapted to other heuristics by changing to an appropriate $f(t_i)$. Although MBTI and Belbin-based team evaluation heuristics have been used in the past to form successful teams in the classroom compared to traditional grouping criteria (i.e., self-assembled, random, staffed teams), they have not been compared with each other in terms of team and class experience. While both may achieve satisfactory results compared to traditional grouping criteria, it is necessary to study which heuristic obtains a better class and team experience, as well as which heuristic can be more easily deployed in a real classroom. This is one of the main gaps identified in the literature.

3.3.1 The Belbin heuristic

Belbin's role taxonomy defines one of the most important theories regarding successful team dynamics. In this theory, Belbin identifies eight behavioral patterns that are present in many successful teams: Implementer (IM), Coordinator (CO), Shaper (SH), Plant (PL), Resource Investigator (RI), Monitor-Evaluator (ME), Teamworker (TW), and Completer-Finisher (CF). There are some later versions of the taxonomy that include a ninth role: the Specialist. However, we did not include this role in our study as all of the courses involved in the experiments are introductory courses. Therefore, we consider that all of the students start in the subject from the same starting point, and there is no specialist per se. However, the presented team evaluation heuristic can be easily adapted to include the ninth role in its calculations.

Belbin's theory describes all of the behavioral patterns that are present in successful teams. Therefore, in theory, teams

should have at least one individual that can exhibit one of the roles. We use the 8-role taxonomy and denote each of the prospective roles as $r_k, k = 1, \ldots, 8$. To use this heuristic, we need to collect the results of the Belbin Self-Perception Inventory [37] for participating students. For each student s_i this inventory produces a numerical score for each of the roles. Let $b_{i,k}$ denote the score received by student s_i at role r_k. Belbin proposed a table (Table 2) that classifies the score obtained by an individual for each of the eight roles according to a salience level: low, average, high, and very high [43].

We assume that a student s_i can exhibit the behavior in a Belbin role as long as he/she obtains a high or very high score in the role. For simplicity, we denote as β_k the threshold needed in the Self-Perception Inventory to achieve at least high in role r_k. Then, we define the team score obtained by team t_i for role r_k as:

$$f_k(t_i) = \begin{cases} 1 & \exists s_j \in t_i, b_{j,k} \geq \beta_k \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

In other words, the team scores a point considering role r_k as long as at least one team member receives a high or very high score in the score associated with the role in the Self-Perception Inventory. Then, the team performance $f(t_i)$ according to the proposed heuristic can be calculated as:

$$f(t_i) = \frac{1}{8} \times \sum_{k=1}^{8} f_k(t_i) \tag{2}$$

As a consequence, the heuristic evaluates to 0 when no team member achieves at least a high score in any role, and it evaluates to 1 when each role can be played by at least one of the team members.

3.3.2 The MBTI heuristic

The Myers-Briggs Type Indicator (MBTI) is an instrument that focuses on identifying an individual's personality in four

Table 2 Thresholds for achieving different scores in each of the Belbin roles according to the Self-perception questionnaire

	Low	Average	High	Very High
Implementer	0-6	7-11	12-16	>16
Coordinator	0-6	7-10	11-13	>13
Shaper	0-8	9-13	14-17	>17
Plant	0-4	5-8	9-12	>12
Resource Investigator	0-6	7-9	10-11	>11
Monitor-Evaluator	0-5	6-9	10-12	>12
Teamworker	0-8	9-12	13-16	>16
Completer-Finisher	0-3	4-6	7-9	>9

different dimensions. Each of these dimensions is formed by a bipolar scale that represents two opposite personality traits: Extroversion (E) - Introversion (I); Sensing (S) - Intuition (N); Thinking (T) - Feeling (F); and Judging (J) - Perceiving (P). The combination of these four dimensions then makes up 16 different personality types.

The basis for team formation in this team evaluation heuristic is focused on the balance of personality types. As pointed out by some studies [4, 24, 27], team performance may be linked to the diversity of personalities found on the team. This indicator suggests that team performance is related to a well-balanced diversity of psychological types. The MBTI-based heuristic used in this article is a normalized version of the heuristic presented in [44]. We use this heuristic as a simple representative of heuristics based on the MBTI indicator, although it is acknowledged that there are also other MBTI-based heuristics in the literature. The heuristic aims to achieve heterogeneity in the team with regard to the different scales used in Myer-Briggs. The MBTI test assigns a different type for each of the four dimensions: the Intuition/Sensing dimension, the Extraversion/Introversion dimension, the Thinking/Feeling dimension, and the Judging/Feeling dimension. As every dimension has exactly two opposite traits, we say that dimension k has two possible assigned traits, referenced as k_1 and k_2, respectively. We define a binary function $\gamma_k(s_j, k_l)$ that returns 1 when student s_j has been assigned trait k_l by the MBTI test in dimension k. Then, the score obtained by a team t_i in a dimension k is defined as:

$$f_k(t_i) = \begin{cases} 0 & \text{if } \exists k_l, \sum_{s_j \in t_i} \gamma_k(s_j, k_l) = \text{t\_i} \\ 1 & \text{if } \exists k_l, \sum_{s_j \in t_i} \gamma_k(s_j, k_l) = 1 \\ 2 & \text{otherwise} \end{cases} \tag{3}$$

Then, if all of the team members have the same trait for a dimension, the team does not receive a score for that dimension. If all but one team member have the same trait for a dimension, then the team receives a score of 1 for that dimension. Finally, if there is a certain diversity of traits for a dimension (at least two members), the team receives a score of 2. Then, the score of a team is calculated as follows:

$$f(t_i) = \frac{1}{8} \times \sum_{k=1}^{4} f_k(t_i) \tag{4}$$

The heuristic evaluates to 1 when there is a variety of traits for every MBTI dimension, and it evaluates to 0 when all of the team members have the same MBTI personality.

4 Experiments

We have used different versions of this tool in the Bachelor's Degree Program in Tourism since the academic course of 2014. Throughout this project, several courses, lecturers, and students have participated in the use of this tool. This has allowed us to carry out different evaluations on the tool. First, the participation of several courses throughout several academic years has allowed us to carry out a long-term experiment to compare the performance of the team allocations provided by the two team evaluation heuristics described in Section 3.3. Second, we have had the involvement of several lecturers throughout our experience, starting in 2014 to this day. Working in the classroom with the teams formed by the tool has allowed the lecturers to form an opinion on the use of these types of tools in the classroom and compare their use to the traditional methodologies that had previously used for team formation. Finally, the collection of students' profiles using Belbin's role taxonomy and MBTI has allowed us to build a descriptive profile on the types of students that participate in the Bachelor's Degree Program in Tourism. We hope that this information helps us in the future to tailor more personalized learning and social activities in the Degree Program.

In this section, we describe the experimental evaluation that we carried out. First, we describe the courses that have participated at some point in the formal or informal use of the computational tool as well as the team activities where students have participated as a team. This information will help the reader understand the kind of activities and courses where this tool has been used. Second, we describe the data collection procedure designed to evaluate the performance of the MBTI and the Belbin heuristic presented in this article. Afterwards, we analyze the MBTI and Belbin roles of the students that participated in the experiments carried out to evaluate the performance of both the Belbin and MBTI heuristics. The goal of this analysis is twofold. On the one hand, the analysis allows us to profile the personality and behavioral patterns of our student cohorts. On the other hand, the results of this analysis are important for understanding the results of the comparison carried out between the Belbin and the MBTI heuristics since the imbalance of roles and personalities may affect the effectiveness of the application of the two heuristics. Then, we analyze the experiment carried out to compare the two heuristics in order to identify the heuristic that can form teams that have better team dynamics and provide a better class experience. We then describe the findings of a formal interview that we conducted with lecturers that participated in the use of the computational tool. The goal of the interview is to qualitatively assess the perception of the lecturers on the impact of the tool, its ease of use as well as its effectiveness compared to traditional team formation. Finally, we provide some discussion and conclusions about

the results and insights that we obtained from the experience and discuss the limitations of our study.

4.1 Description of activities

As mentioned, we carried out experiments by using our tool throughout several academic years to assess its use and performance in classroom environments. More specifically, we have collected data from its use in different semesters of the Bachelor's Degree Program in Tourism from the academic year 2014/2015 to 2018/2019. The course 2019/2020 is not included in the study due to the special circumstances of COVID and how it affected the normal class activities. During these years, we have had the collaboration of several lecturers and courses that participated in our experience. The courses involved are courses that have team-based coursework. Following, we provide a short description of the courses (see Table 3) as well as the team activities for each of the courses involved.

As the reader may observe, the *Catering Production Management*, *New Technologies Applied to Tourism*, and *Business English* courses share the same description for the team activity. The reason behind this is that the students participate in an integrated project where they apply the competencies and skills learned throughout the three courses. In all of the courses, the students participate in projects that span several weeks during the semester. Therefore, the data that we

collected about the tool usage and the performance of both the Belbin and MBTI heuristics focuses on project-based activities. It is acknowledged that the results may differ for other types of team activities that are carried out in the classroom.

4.2 Data collection

The goal of this section is to describe the data collection procedure and the experiments that we carried out in order to identify which of the two team formation heuristics may lead to the most satisfactory results. With that goal in mind, we devised the following experiment.

Each of the aforementioned academic years, we carried out a data collection process to obtain the Belbin and MBTI profiles of students involved in the described courses. More specifically, at the start of the academic year, the students were informed that they could participate in a research study related to identifying successful team formation strategies in the classroom. We decided to focus on the third-year courses of our Bachelor's Degree Program in Tourism, where two long-term team projects are used as assessment tools: one in the first semester, and another one in the second semester. For each academic year, the students were told that they would complete two questionnaires at the start of the course to assess their Belbin role and MBTI personality. Throughout this experience, a total of 260 students decided to participate in the study and completed both the Belbin self-

Table 3 Description of the courses

Course: Tourist Territorial Resources	Semester	2A
Activity	A project to analyse the climate and tourism of a given region	
Learning outcomes	-To work with climate as a resource for tourism guidance	
	-To establish climate impact in tourism and tourist activities	
Course: Travel management	Semester	3A
Activity	A project of travel management	
Learning outcomes	-To design a creative package travel	
Course: Catering Production Management	Semester	3B
Activity	An entrepreneurship project and appropriate landing page	
Learning outcomes	-To develop creativity techniques	
	-To apply technology in projects	
	-To integrate catering knowledge in real applications	
Course: New Technologies Applied To Tourism	Semester	3B
Activity	An entrepreneurship project and appropriate landing page	
Learning outcomes	-To use programming languages for web development	
	-To design web pages	
Course: Business English	Semester	3B
Activity	An entrepreneurship project and appropriate landing page	
Learning outcomes	-To use persuasive and marketing language	
	-To develop oral skills	

perception inventory and the MBTI personality test. Of those 260 students, 162 are female and 98 are male students. Therefore, almost two-thirds of our sample corresponds to female students. This proportion is consistent with the enrollment statistics in our Bachelor's Degree Program in Tourism.

Then, for the duration of the project course, they would be allocated to a team based on some criteria unknown to them. We randomly split the group of students into two halves and applied the Belbin heuristic strategy to one half and the MBTI heuristic to the other half. In the following semester, we exchanged the application of the MBTI and Belbin heuristic for each of the halves. Therefore, each half was grouped by both the Belbin and the MBTI heuristic. In those experiments, all of the teams formed had four members, except for some teams with five members. The lecturers used the computational tool to form the teams in the two semesters using the assigned criteria. The students were not informed of the specific criteria employed to form teams in order to avoid bias.

After each team project, at the end of the semester, a satisfaction form was delivered to the students to gather their perceptions about the experience of working with their teams. To assess the quality of the team experience, we designed a questionnaire that covers two areas: student satisfaction and team dynamics. Specifically, we designed a questionnaire with seven different questions that are categorized and described in Table 6. All of the questions are presented on 5-point Likert scales. We collected a total of 175 responses in post-activity questionnaires from the participants, with 66 responses associated with the participants on the teams formed by the MBTI strategy, and 109 responses for the participants on the teams formed by the Belbin strategy. The difference in the number of participants corresponds to different class sizes (e.g., repeating students, exchange students) as well as some participants who did not answered the final surveys. Consent was collected for those students that decided to participate in the study.

4.3 Class profile

In this section, we make an exploratory analysis of the class profile related to the Belbin and MBTI profiles. Following, we describe the most relevant findings in our student population concerning the MBTI personality test.

We calculated the most salient personality type for each of the four dimensions of the MBTI test: Extraversion/Introversion, Sensing/Intuition, Thinking/Feeling, and Judging/Perceiving. The pie charts shown in Fig. 2 briefly describe our student sample distribution for each of the four dimensions. As the reader may observe, most dimensions seem balanced except for the Extraversion/Introversion dimension, where we have a majority of extroverts. The implications for the MBTI grouping strategy are that, since

dimensions tend to be balanced, it may be easier to form heterogeneous teams than in the case of highly unbalanced distributions.

We decided to delve further into our sample,k and we studied the distribution of the MBTI dimensions based on the gender of the student. We built contingency tables for each MBTI dimension and gender, and we carried out a chi-squared test of independence to assess whether we had enough evidence to discard independence in our population between gender and each MBTI dimension. The null hypothesis (i.e., independence) was rejected for the contingency table representing the distribution of Feeling/Thinking based on gender ($p - value = 0.001$, $\alpha = 0.05$). Therefore, we concluded that there is evidence to think that female students tend to have a more salient feeling personality type than male students. Specifically, the sample proportions were approximately 62% Feeling/38% Thinking for our female students and 41% Feeling/59% Thinking for our male students. Other dimensions showed similar proportions for both male and female students, and the null hypothesis cannot be rejected. When jointly analyzing the four dimensions, we also found significant differences between the personality distribution among male and female students. We carried out a Fisher exact test on the contingency table formed by MBTI personalities and gender, and the null hypothesis of independence was rejected ($p - value = 0.004$, $\alpha = 0.05$). Therefore, our female and male student populations have different personalities. Table 4 shows the specific distribution of MBTI personalities in our student sample. As can be observed, there are clear differences in the distribution of roles based on gender. This is particularly true for those roles that involve the Thinking personality, which tends to be more frequent in the male student population. This, of course, is in line with our previous analysis, where we identified that the Thinking dimension tends to be more salient among male students than
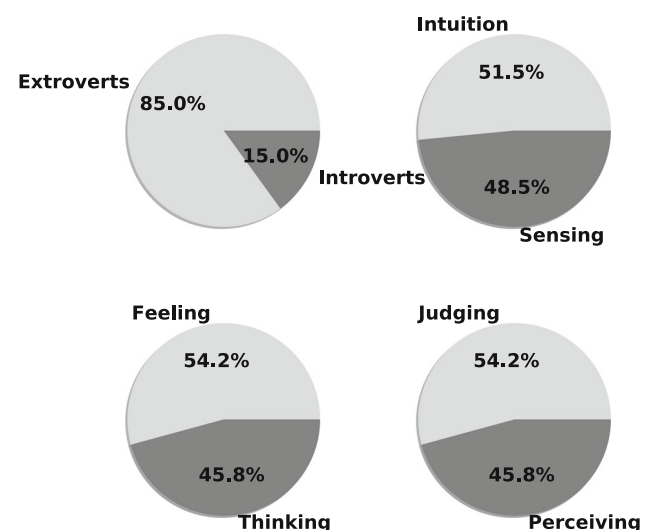

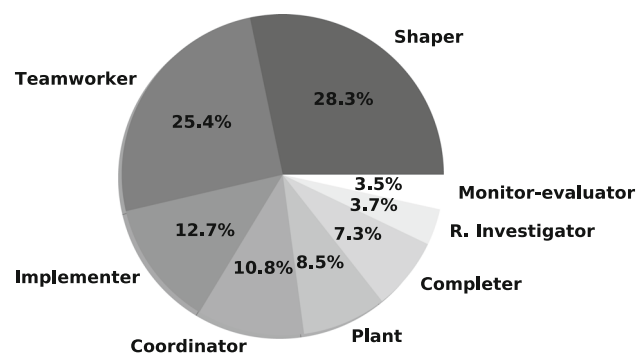**Fig. 2** Most salient MBTI dimensions for our student sample

Table 4 Distribution of MBTI personalities (percentage) among male and female students in our sample

	ENFJ	ENFP	ENTJ	ENTP	ESFJ	ESFP	ESTJ	ESTP
Male	24.07	14.20	5.00	4.32	14.82	2.47	21.00	0.62
Fem.	10.20	6.12	14.28	11.22	11.22	8.17	20.40	1.02
	INFJ	INFP	INTJ	INTP	ISFJ	ISFP	ISTJ	ISTP
Male	0.62	2.47	1.85	0.62	3.00	0.62	4.32	0.00
Fem.	4.10	0.0	3.07	0.00	1.02	0.00	9.18	0.00

it is for female students, while the Feeling dimension tends to be more salient among female students.

We conducted a similar analysis for the most salient Belbin roles in our population. We calculated the two most prominent Belbin roles for each of our students, and then we analyzed the frequency of each role in our sample. As expected, the sample data suggests some roles are more frequent than others. Figure 3 shows how roles are distributed among the set of the two most prominent roles of each student. As one can observe, the two most common roles in our sample are Shaper and Teamworker, which account for more than half of the sample. The next more frequent roles are Implementer and Coordinator, but they are just found in approximately one-fourth of the sample. This role heterogeneity makes it difficult to create fully balanced teams, as some roles are more infrequent and, therefore, will most likely not be included in many teams. This may be of special concern for the Belbin-inspired heuristics since Belbin's theory suggests that all roles should be present in order to have successful teams.

We also studied any potential gender differences in our population with regard to the most prominent Belbin role. We carried out Mann-Whitney tests on the scores obtained by each gender for each Belbin role dimension in the Self-Perception questionnaire. The tests suggested that male students tend to score higher than female students in our stu-

dent population for the *Coordinator* dimension ($p-value = 0.03$, $\alpha = 0.05$), female students tend to score higher in the *Shaper* dimension ($p-value = 0.002$, $\alpha = 0.05$), and male students tend to score higher in the *Monitor-Evaluator* dimension ($p-value = 0.0005$, $\alpha = 0.05$). In line with these results, a Fisher exact test on the contingency table of both gender and the most prominent Belbin role of each student suggests that there are significant ($p-value = 0.004$, $\alpha = 0.05$) gender differences in our population concerning the distribution of some roles. One may observe in Table 5 that the Shaper role is more frequent in females than it is in males, that Monitor-Evaluator role was never the most prominent role for our female students, and that there is some marginal difference between the frequency of the Coordinator role between male and female students. We also observed that the Teamworker role was the most predominant role for male students, more frequently than for female students.

This analysis provides some insights. First of all, there is a clear imbalance in the distribution of Belbin roles in our student population, with some behavioral patterns like Monitor-Evaluator, Resource Investigator, Completer, and Coordinator being more unlikely than a uniform distribution of the roles. According to Belbin's role theory, these behavioral patterns or roles should be present in successful teams. Hence, students should be prepared to play these roles more naturally when needed. It was of particular concern to us that the data collected from our student population indicated that there may be gender differences in the distribution of some roles. For instance, the coordinator role was only found as the primary role for female students for 4.9% of the sample, while it almost doubled that frequency in the case of the male sample. The coordinator role is typically associated

Fig. 3 Frequency of each Belbin role in the two most prominent roles of each student

Table 5 Main Belbin role frequency (percentage) among female and male students

	CF	CO	IM	ME	PL	RI	SH	TW
Male	9.1	9.1	6.1	4.1	6.1	1.0	26.5	37.7
Fem.	3.0	4.9	6.7	0.0	8.6	3.7	43.3	29.6

with leadership and management skills. We found this issue concerning, as our goal as an institution is to train professionals for an equitable society. We believe that the cause for this issue may be associated with some lingering cultural and societal bias. In general, this shows the need for empowering our students' management and leadership skills, particularly in our female student population.

Another insight to take from this analysis is the imbalance present in both the MBTI personality dimensions and the Belbin roles. However, the imbalance is particularly notable in the Belbin roles, as mentioned. This may make it difficult to form balanced teams in our current student population, which may favor the use of one heuristic with respect to the other.

4.4 Comparison of team evaluation heuristics

In this section, we analyze the in-class performance of the two heuristics proposed in this article. In order to achieve this goal, we analyze the 175 responses collected from the post-activity questionnaires described in Table 6. As the reader may remember, the questionnaire is divided into two sections: student satisfaction and team dynamics. Following, we analyze and compare the responses collected from both parts of the questionnaire.

4.4.1 Student satisfaction

First, we start analyzing those questions relating to student satisfaction with the team experience. To analyze data from Likert scales, we combined categories into binary categories (i.e., positive and indifferent/negative categories) since some of the options did not have enough samples to generalize (i.e., less than five counts). When combined into binary categories, the result for each question is a 2×2 contingency table. 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 [23, 25]. 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. In order to analyze the data in contingency tables, we employ a test for independence using the mid-p method as carried out in the life sciences [47, 55].

The detailed results for the experiment can be found in Table 6 and Fig. 4. The table shows the questionnaire's content and the percentage of positive responses collected for each team evaluation heuristic, as well as the mid p-value associated to the independence test carried out on the 2x2 contingency tables. The gray shading shows the heuristic that obtained statistically better results for the question. As can be observed, both heuristics tend to obtain positive results, as sustained by the literature [1, 3, 6, 7, 31, 36, 50, 62], but the MBTI heuristic tends to collect a higher percentage of positive responses and is never worse than the Belbin-based heuristic. The figure shows the detailed responses of the students for questions where we found statistical differences. This figure shows how, overall, there is a tendency for individual to perceive the experience more positively if they have been involved in a team formed by the MBTI heuristic.

We now analyze in depth the questions concerning student satisfaction. Question Q1 assesses the general satisfaction of the students with their experience on their respective teams. The null hypothesis of the test is that the team evaluation heuristic does not have any effect on the proportion of positive responses, and the alternative hypothesis is that one of the grouping strategies is more likely to produce a positive response. The test produced a mid p-value of 0.047

Table 6 Contents and results of the post-activity questionnaire

Student satisfaction				
ID	Question	MBTI	Belbin	Mid p-value
Q1	How would you assess your experience on this team?	78.7%	66.9%	0.047
Q4	I am satisfied with the work carried out in this project	86.3%	85.3%	0.43
Q7	I would work again with X (X being a team member)	87.8%	76.14%	0.028
Team dynamics				
Q2	There were clear norms and tasks for everyone	93.9%	82.5%	0.014
Q3	Each team member did his/her part of the work	83.3%	73.3%	0.06
Q5	Team members accepted criticism in a positive way	87.8%	82.5%	0.17
Q6	It was easy to make decisions as a team	87.7%	75.2%	0.03

The results show the percentage of positive responses for MBTI and Belbin, as the mid p-values for the statistical test
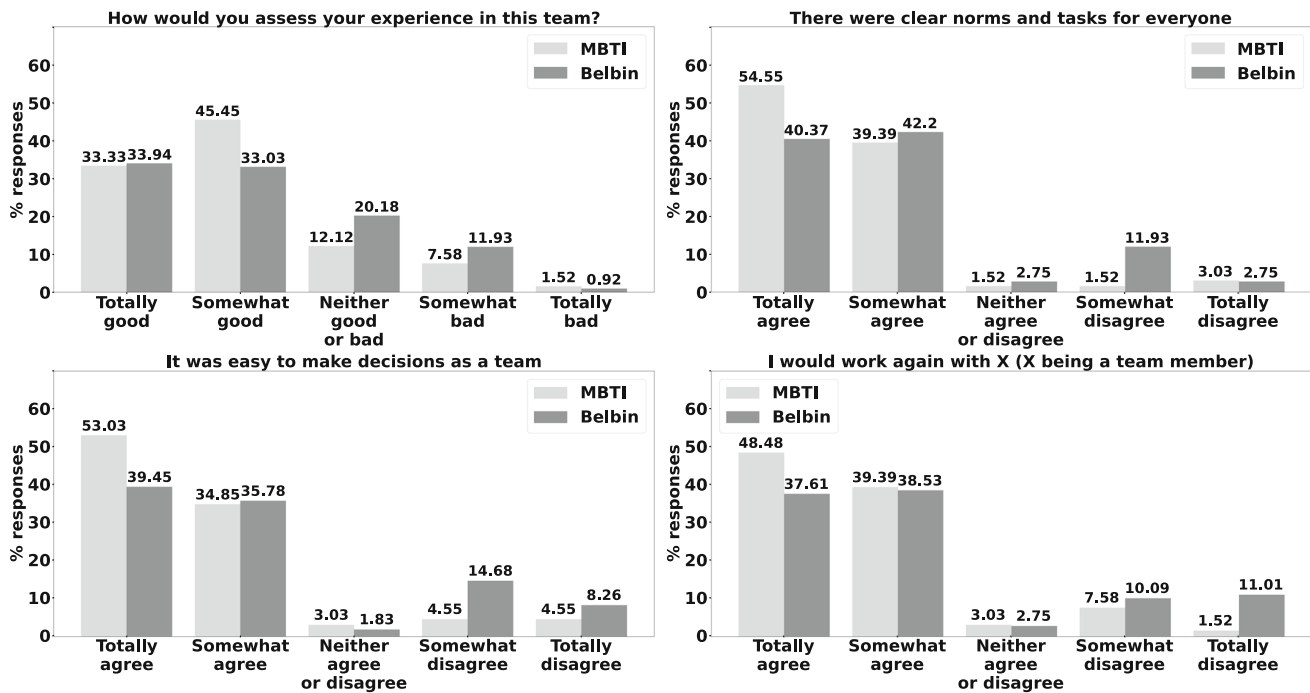
Fig. 4 Detailed responses to the post-activity questionnaire for questions Q1, Q2, Q6, and Q7

($\alpha = 0.05$), which suggests **support for the students being more satisfied with the general team experience when using the MBTI grouping strategy**. As Table 6 and Fig. 4 shows, 78.7% of the participants provided a positive response (i.e., Totally good or Somewhat good) for the MBTI heuristic, while only 66.9% of the participants provided that response for the Belbin heuristic, with more students being indifferent or providing a negative answer in the Belbin setting. The same applied to Q7, which asked the students about the desirability of working with each of their teammates. The mid p-value obtained for the odds ratio test was 0.028 suggesting **the support for the MBTI teams is more likely to produce positive evaluations of their team members than the participants on Belbin teams**. The percentage of students that provided a positive response was 87.8% and 76.15% for the MBTI and the Belbin heuristics, respectively. In addition, as observed in Fig. 4, the students involved in the MBTI experience lean towards totally agreeing with the statement, while the students in the Belbin experience are evenly divided between totally and somewhat agreeing. When analyzing the satisfaction of the students with the contents of the project that they carried out, we did not find strong evidence for either of the two grouping strategies being more likely to more positively perceived the quality of the project that was carried out (mid p-value of 0.43). In both grouping strategies, most of the participants in the MBTI and Belbin groups positively perceived the work that they carried out, with percentages of 86.3% and 85.3%, respectively. Both findings **may indi-**

cate that the MBTI grouping strategy may produce more positive student satisfaction with regard to the team experience than the Belbin grouping strategy, although it may not particularly affect the perception of the project carried out. There is no evidence for students more positively perceiving the work that they carried out in one of the two grouping strategies (Q4), although both provide high satisfaction.

4.4.2 Team dynamics

As stated in Table 6, there are a total of four questions associated with team dynamics. Q2 asks the students about the existence of clear norms and tasks for everyone. Therefore, the question is related to team coordination. After carrying out the test, the mid p-value is 0.014, **which supports the alternative hypothesis of MBTI teams perceiving more clear norms and task distribution**. The percentage of positive responses stands at 93.9% for the MBTI heuristic and at 82.5% for the Belbin heuristic. The students grouped by the MBTI heuristic tend to totally agree with the statement, while a higher percentage of the students show indifference or a negative response in the Belbin setting. Q3 is related to each team member fulfilling the tasks that he/she was assigned. After carrying out the odds-ratio test on the 2x2 contingency table, we obtained a mid p-value of 0.06, which is not enough evidence to reject the null hypothesis of the odds ratio being equal to 1. Therefore, we cannot conclude

that either grouping strategy resulted in more commitment to the task distribution. However, a look at the percentage of positive responses for each grouping strategy reveals that 83% of the students in the MBTI grouping felt that members fulfilled their tasks, while only 73% of the students felt that way in the Belbin grouping. This may suggest a difference, although a larger sample size may be necessary to clarify this. As for Q5, which is related to how team members accept criticism and others' opinions, we did not find specific support for either of the two grouping mechanisms being more likely to produce a positive response in students. Both obtained a high percentage of positive responses: 87% and 82% for MBTI and Belbin, respectively. Although there is a 5% difference in the sample, the mid p-value of 0.17 does not provide enough evidence to reject the null hypothesis. Q6 is related to team decision-making, and the test carried out **suggests that it is more likely for the students in the MBTI grouping to feel more positive about the decision-making processes carried out on the team compared to the students grouped using the Belbin strategy** (mid p-value of 0.03, $\alpha = 0.05$). In this case, 87.7% of the students in the MBTI setting provided a positive response, while 75.2% of the students in the Belbin setting provided a similar response. A more in-depth look at Fig. 4 shows that, again, the students grouped by the MBTI heuristic tend to totally agree with the statement, while more students in the Belbin setting tend to show indifference or negative responses. For two of the questions related to team dynamics, the tests carried out revealed that positive responses were more likely to be elicited from the MBTI grouping strategy than the Belbin strategies. This is especially the case for the task and norms organization and for the decision-making processes. The results obtained may suggest that the students were more likely to perceive that each team member did his/her part of the work in the MBTI grouping, although a larger sample size may be needed to confirm this. We did not find any evidence in any item for higher positive responses for the Belbin strategy compared to the MBTI strategy.

4.5 Interview with the lecturers

After the computational tool was used by the lecturers for several years, personal interviews were carried out with the six lecturers involved throughout the years in the use of the computational tool in the classroom. This sample size corresponds to all of the lecturers that participated throughout their courses in the designed experiment. The goal of these interviews was to obtain a qualitative evaluation of the experience and the computational tool. Following, we summarise the main conclusions that we extracted from the interviews, providing some of the specific comments that the lecturers gave us.

First, we asked the lecturers about the students' satisfaction with working on the teams formed by the tool. At the beginning of the experience, the lecturers found a general rejection by the students since the majority of them preferred to work on self-assembled teams rather than being obligated to work with specific teammates. After some years of using the team formation tool, **this system has become more widely accepted**, although some students would still prefer to choose their own teammates to work with. Even though it cannot be generalized, the lecturers feel that **the students with higher expectations are those that usually show more rejection** to working with teammates that are not chosen by them. Actually, these students would even prefer to work individually rather than work with others. The lecturers think that this may be because these students think that they could achieve higher marks by working individually. In contrast, exchange students and those students that are less sociable are usually more willing to work with teammates that are chosen by the tool. This can be explained due to the fact that these students are no longer responsible for finding teammates to work with. In general, the lecturers think that **explaining to students that teams may be formed by following some scientific criteria can be very positive** for the acceptance of team formation. In addition, it is very important to provide them with arguments to strengthen the benefits of teamwork. Following, we quote one of the lecturers' opinions on this issue.

> **Answer:** *"At the beginning of the semester, I suggest explaining to students the objective of teamwork and why they would not end up choosing their teammates in professional scenarios. Over the years, we have figured out that providing this information is very important to avoid rejection by students. At the end of the semester, students recognize (sometimes surprisingly) that the experience was very positive and the results were satisfactory."*

With regard to complaints, the lecturers state that according to their own experience, when students work on teams that are formed by them, there are almost no complaints. In contrast, when working with teams formed by the tool, students usually complain when they are grouped with students that they consider to have low performance. This is especially critical when students with very high expectations are forced to work with other students whose expectations are very different. This issue could be addressed by including some peer-assessment in order to weigh the contribution of each team member. Actually, some of the lecturers interviewed already include this type of evaluation. Even though the final mark, they should not feel teamwork is a threat. In spite of these complaints, some of the lecturers claim that teams whose members are continuously complaining are usu-

ally able to obtain very good results. What is more, **at the end of the project students who previously had prejudices for working with other specific students stated that the teamwork experience was very positive**.

We also asked the lecturers about the development of teamwork-related skills. In general, lecturers think that the computational tool **provides students with a scenario that is very similar to what they could find in the real world**, i.e., the need for working on heterogeneous teams with people that are not chosen by them to develop a project. To achieve this goal, at the beginning of the project, each team must agree to some commitments: what the strengths and weaknesses of each member are, when and how meetings will take place, how tasks will be divided and scheduled, how delays will be detected and addressed, how conflicts will be solved if they occur, and so forth. The lecturers think that **these skills are developed in more depth when teams are formed by a tool** since, when students select their friends, they usually protect the lack of work of friends, mitigating delays, conflicts, or poorly developed tasks. Below, we present some quotes made by the lecturers at the interview.

> **Answer:** *"When students work on a team, it does not mean that they develop teamwork skills. This tool provides a great opportunity to develop teamwork skills since students work with other students that are not chosen by them, which is similar to what they may find in a real-world situation."*

> **Answer:** *"I would remark two key points that students learn. First, they have to talk about their skills to better organize the tasks. Second, they learn about conflict resolution. Conflicts appear in almost every team. We give students some guidelines to solve conflicts, but they have the responsibility to talk and reach agreements."*

> **Answer:** *"It is essential that students learn to work with any type of people beyond their close group of friends."*

With regard to the initial questionnaires (i.e., Belbin and MBTI), some lecturers think that students are not fully motivated to complete them. Some lecturers also remarked that, despite this lack of motivation, what motivates most students is knowing the results of the tests (i.e., roles and personalities). According to the lecturers, one of the most positive aspects of using the team formation tool is that it allows the **creation of balanced teams** with different profiles. In addition, the experience of working with different people is also well-valued by lecturers. Below, we quote some of the lecturers' opinions.

> **Answer:** *"Every year, I am more convinced that the teams formed by the tool are very productive. It also has*

an impact on learning organizational and professional skills."

> **Answer:** *"The key is to get the students out of their comfort zone. This gives them security for their professional career. After this experience, they are not scared anymore of working with people whom they did not know previously."*

As negative aspects, some lecturers mentioned that it may be desirable to take other factors into account in the formation. These factors could be related to individual expectations or academic results. In contrast, other lecturers do not agree with this, since this would create imbalanced teams (teams with very good performance and teams with very bad performance). These requirements might be taken into account easily by incorporating new team evaluation heuristics, something that our optimization model allows for. Following, we present some of the lecturers' opinions on this issue.

> **Answer:** *"In work environments, one may have to work with colleagues with different expectations, and students have to learn how to deal with it. The development of teamwork skills does not only depend on the final result."*

In conclusion, all of the lecturers considered this team formation tool to be very useful for teamwork in academic environments and that it should be extended to other courses. It should be also highlighted that none of the lecturers interviewed knew of any other tool for team formation and management in academic environments. Finally, we show the opinion of one of the lecturers with regard to the use of the tool.

> **Answer:** *"I would strongly recommend the use of this tool to other lecturers that use teamwork. Absolutely."*

4.6 Discussion

We would like to start this discussion by remarking that, in general, the experience has been quite positive. First of all, this experience has allowed us to study our student population in depth. We have observed how some behavioral patterns in teams are very rare in our student population. This suggests that we need to foster some behaviors in our students' teams, as students should learn how to play diverse roles on teams. To do this, we are planning to design specific course activities and workshops. There have also been some concerns with regard to some gender differences observed with some of the roles, which are probably biased by cultural and societal issues. This suggests that we should foster and empower the leadership and management skills of our female student population.

Second, we have compared the performance of two heuristics for team formation: one based on Belbin's theory, and another based on MBTI. It should be stated that both grouping strategies achieved a high number of positive responses. Both theories have been reported in several works as being useful for the formation of teams in the classroom [1, 3, 5, 7, 13, 18, 63, 65]. However, the experiments that we carried out suggest that the MBTI grouping strategy results in students being more likely to feel positive about both the team experience and team dynamics, i.e. the heuristic was better than or equal to the Belbin heuristic in all aspects. Of course, the results should be analyzed with caution.

It should be highlighted that the success of these grouping strategies relies on the ability to form heterogeneous teams in the classroom. As observed in the results from Section 4.3, it may be easier to form heterogeneous teams in the MBTI setting, as MBTI dimensions seem to be balanced (i.e., around 50-50% split) for all of the dimensions except for introversion/extroversion. However, when analyzing Belbin's role distribution, it is observable that some roles are much more frequent than others. Belbin's theory states that all of the observed roles should be present on a successful team. Fewer roles may mean less effective teams. While a company or organization may have flexibility in forming teams, considering that they may hire roles that they lack, this is not the case in the classroom setting. Therefore, it may be more difficult to form well-balanced teams following Belbin's theory. If one assumes both theories to be equally effective, this may be translated as a slight disadvantage when applied to the classroom setting due to the prospective Belbin role imbalance. Figure 5 precisely illustrates this issue. The figure shows the cumulative distribution plot for the team evaluation heuristic (scaled from 0 to 1) of teams formed with both the MBTI and the Belbin grouping strategy. As one may observe, there is a tendency for Belbin teams to score lower, as only approximately 49% of the teams formed with this strategy achieved more than half the highest possible metric value. This percentage was reduced to 15% when considering teams scoring higher than 60%, the highest possible metric value. Also, 63% of the teams formed with the MBTI heuristic achieved more than half the highest possible metric value, while 40% of the MBTI teams scored higher than 60% of the highest possible metric value. The differences between the two cumulative distribution plots illustrate how forming heterogeneous Belbin teams may be more difficult than forming heterogeneous MBTI teams in a classroom setting. This may help to explain the differences in student satisfaction and perceived team dynamics, as observed in the previous section. Hence, one can conclude that there were differences in applying both the MBTI and the Belbin team evaluation heuristics in the classroom, but those differences may be explainable due to the classroom profile and not necessarily due to the validity of the theories. In our experience, with regard to our

student population and assuming a lecturer role, it may be more positive to apply MBTI grouping strategies than it is to apply Belbin grouping strategies.

In addition to the previous findings, we have gained some interesting insights from the interviews conducted with the lecturers responsible for using the tool in the classroom throughout these years. In general, there is an initial objection to working as a part of a team that is not formed by the students themselves. This is more acute in high performing students when placed on a team with low performing students. Nevertheless, as pointed by the lecturers, this initial rejection can be tackled by making students understand that real-world teams are never chosen and that one should learn to work as a team with any individual. In fact, the lecturers suggested that some of the critical students valued the experience very positively by the end of their projects. This general positive feeling is also supported by the student satisfaction questions in the surveys conducted to evaluate the two heuristics. Moreover, the lecturers also believe that the creation of balanced teams allows students to develop real teamwork skills, while self-selected teams do not since the scenario is more similar to the real-world. The team dynamics and student satisfaction when using team formation tools has also shown to be of higher quality than randomly selected teams, as pointed out by previous studies [1, 5, 6].

Finally, the authors of this article would also like to point out some limitations of this study. First of all, the results of this experience apply to our student profile (i.e., Tourism
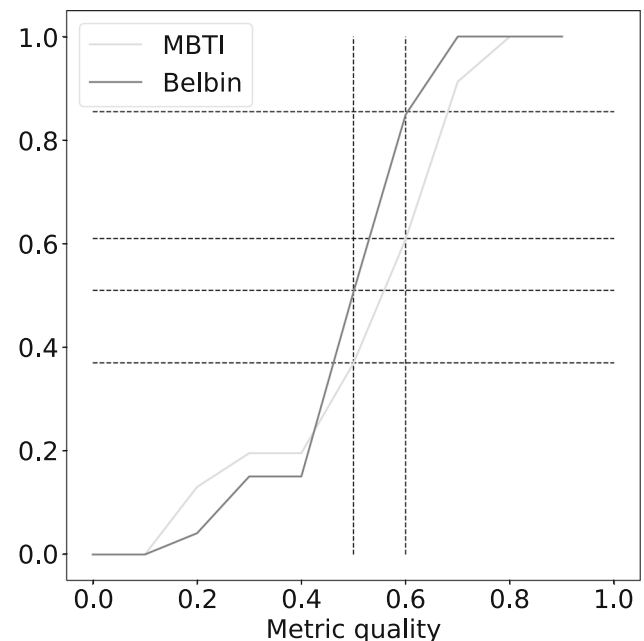

Fig. 5 Cumulative distribution plot for the metric of teams formed by the MBTI and the Belbin heuristics

students). Other classroom profiles may be different and, perhaps, more unbalanced with respect to the MBTI dimensions. In those situations, it may be more positive to apply a Belbin grouping strategy. Despite this, there are general conclusions that can be drawn for other contexts. The authors of this article recommend that lecturers carry out a pre-analysis to assess the heterogeneity of the classroom since it may be linked to the final experience of students on teams. A previous analysis may indicate which team formation heuristics may work better in practice, rather than applying team formation heuristics as black-boxes, which is something common in the literature. In addition to this, the type of activities in which our students participated are long-term projects. We acknowledge that the results may be different for short-term team activities, as team dynamics may not develop.

5 Conclusions

Team formation is a major issue when dealing with teams in the classroom because students should learn to work as a team and not have negative experiences with teamwork that precludes them from being willing to work as a team. As a response to this, the team formation problem has been studied from different perspectives in computer science. One of the strands in computer science is designing algorithms and tools for the automatic formation of optimized teams. As a result, several team formation algorithms and tools, backed by different theories, have been proposed for team formation in the classroom. At the core of the construction of optimized teams is the team evaluation heuristic. This is just an approximation of team performance since team performance is a complex and multifaceted phenomenon that is difficult to quantitatively assess and that cannot be predicted with precision prior to the task at hand. While some of these team evaluation heuristics have been compared with classic team formation strategies such as random, self-assembled, or staffed-formed teams [1, 3, 6, 7, 18, 31, 34, 36, 48, 50, 53], there is little research on the effectiveness of the MBTI and the Belbin heuristics when compared with each other. In addition to this, many of the studies found in the literature focus on the application of team formation tools, but they do not provide insights on the issues and the benefits that may arise from the experience of applying this kind of technology in the classroom. In this article, we have presented a team formation tool that implements two team evaluation heuristics based on two popular criteria found in the literature: Belbin's role taxonomy and the Myers-Briggs Type Indicator (MBTI). As we stated, there was a lack of studies focused on comparing different team evaluation heuristics. This work is a first step towards filling in that gap. More specifically,

we conducted an experience with our tool for five academic years in different courses of the Bachelor's Degree Program in Tourism. This long-term experience has enabled us to both compare the performance of the two team evaluation heuristics and to collect the experiences, problems, and insights of lecturers that have applied this type of technology in the classroom over a long period time. In addition to this, we have presented an integer linear programming model that can easily incorporate and extend any team evaluation heuristic while taking advantage of all of the rest of the tool.

Applying both heuristics in a real classroom experience has also allowed us to study our student population, with some surprising insights. According to our analysis, we found significant differences in the distribution of personalities based on gender. Specifically, the thinking dimension is more frequent in male students, while the feeling dimension is more frequent in female students. Similarly, significant differences also appear regarding the Belbin roles. While male students tend to score higher in coordinator and monitor-evaluator roles, female students are prone to the score higher the shaper role. This latter insight has raised some concerns that we should tackle in the future regarding what skills and roles should be fostered in our student population in order to achieve a fair society and fight societal and cultural issues.

With regard to the performance of the two team evaluation heuristics, we found that those teams formed by following the MBTI heuristic were slightly more satisfied than those formed by following the Belbin heuristic. This is also related to the team experience, which showed that the MBTI teams were more positive regarding the norms, task distribution, and decision-making processes. These results could be due to the fact that it was easier to form heterogeneous teams by following the MBTI criteria in our population, as the dimensions were quite well balanced. In contrast, some Belbin roles were more popular than others, which increased the difficulty of forming teams where the maximum number of roles are present in our student population. In spite of these few differences, we must point out that both strategies had positive responses. Another insight to take from this study is the recommendation to carry out analyzes before applying any team formation algorithm, as they may provide insights on which team evaluation heuristics may work better for one's student population.

We carried out personal interviews with the lecturers involved in the experiments. The interviews allowed us to qualitatively assess the experience of applying this tool in the classroom, to identify the problems that may appear, and to determine what strategies may be more appropriate to manage this type of experience. In general, the lecturers thought that the tool, which provides functionalities for team formation, is very useful for the development of teamwork-

related skills, such as task organization or conflict resolution. According to them, the teams that are formed allow students to focus on situations that may be similar to the ones found in the workplace. Lecturers also consider that the experience for most of the students was beneficial for their education.

As future work, we plan to extend these experiments to other courses in order to observe similarities and differences in the students' profiles. In addition, we are also considering offering the computer-based tool as a standalone application for other lecturers outside of our University.

Declarations

Competing interests The authors did not receive support from any organization for the submitted work. The authors have no competing interests to declare that are relevant to the content of this article.

Ethical and informed consent for data used The study was approved by the Ethical Committee of Universitat Politècnica de Valencia. Informed consent was obtained from all participants prior to their inclusion in the study. Participants were provided with detailed information about the purpose of the study, the procedures involved, and their rights as participants, including the right to withdraw from the study at any time without penalty. Participants were also informed about the measures taken to ensure the confidentiality and privacy of their data.

References

1. Alberola JM, Del Val E, Sanchez-Anguix V, Palomares A, Teruel MD (2016) An artificial intelligence tool for heterogeneous team formation in the classroom. Knowl-Based Syst 101:1–14
2. Alexander PM (2006) Virtual teamwork in very large undergraduate classes. Comput Educ 47(2):127–147
3. Amato CH, Amato LH (2005) Enhancing student team effectiveness: Application of myers-briggs personality assessment in business courses. J Market Educ 27(1):41–51
4. Andrejczuk E, Bistaffa F, Blum C, Rodríguez-Aguilar JA, Sierra C (2019) Synergistic team composition: A computational approach to foster diversity in teams. Knowl-Based Syst 182:104799
5. Andrejczuk E, Rodríguez-Aguilar JA, Roig C, Sierra C (2017) Synergistic team composition. Proceedings of the 16th conference on autonomous agents and multiagent systems pp. 1463-1465
6. Andrejczuk E, Rodríguez-Aguilar JA, Sierra C, Roig C, Parejo-Romero Y (2018) Don't leave anyone behind: Achieving team performance through diversity. 2018 ieee frontiers in education conference (fie) pp. 1-9
7. Aranzabal A, Epelde E, Artetxe M (2022) Team formation on the basis of belbin's roles to enhance students' performance in project based learning. Educ Chem Eng 38:22–37
8. Behfar K, Friedman R, Brett JM (2008) The team negotiation challenge: Defining and managing the internal challenges of negotiating teams. Iacm 21st annual conference paper
9. Belbin RM (2012) Team roles at work. Routledge
10. Borges S, Mizoguchi R, Bittencourt II, Isotani S (2017) Group formation in cscl: A review of the state of the art. Higher education for all, Researcher links workshop, pp 71–88
11. Bravo R, Catalán S, Pina JM (2019) Analysing teamwork in higher education: An empirical study on the antecedents and consequences of team cohesiveness. Stud High Educ 44(7):1153–1165
12. Casper WC (2017) Teaching beyond the topic teaching teamwork skills in higher education. J High Educ Theory Pract 17(6):53–63
13. Chen S-J, Lin L (2004) Modeling team member characteristics for the formation of a multifunctional team in concurrent engineering. IEEE Trans Eng Manage 51(2):111–124
14. Corsino M, Giuri P, Torrisi S (2019) Technology spin-offs: teamwork, autonomy, and the exploitation of business opportunities. J Technol Transf 44(5):1603–1637
15. Farhangian M, Purvis M, Purvis M, Savarimuthu BTR (2016) Modeling team formation in self-assembling software development teams. Proceedings of the 2016 international conference on autonomous agents & multiagent systems pp. 1319-1320
16. Fathi M, Ghobakhloo M, Syberfeldt A (2019) An interpretive structural modeling of teamwork training in higher education. Educ Sci 9(1):16
17. Flores-Parra J-M, Casta?ón-Puga M, Evans RD, Rosales-Cisneros R, Gaxiola-Pacheco C (2018) Towards team formation using belbin role types and a social networks analysis approach. 2018 ieee technology and engineering management conference (temscon) pp. 1-6
18. Flores Ureba S, Simón de Blas C, Borrás-Gené O, Macéas-Guillén A (2022) Analyzing the influence of belbin's roles on the quality of collaborative learning for the study of business fundamentals. Educ Sci 12(9):594
19. Garshasbi S, Mohammadi Y, Graf S, Garshasbi S, Shen J (2019) Optimal learning group formation: A multi-objective heuristic search strategy for enhancing inter-group homogeneity and intra-group heterogeneity. Expert Syst Appl 118:506–521

20. Ge Z, Qi Z, Yang L (2018) A literature review of grouping solutions in collaborative learning. 2018 ieee international conference on progress in informatics and computing (pic) pp. 393-397

21. Gomes AB, Silva M, Valadares DCG, Perkusich M, Albuquerque D, Almeida HO, Perkusich A (2020) Evaluating the relationship of personality and teamwork quality in the context of agile software development. Seke pp. 311-316

22. Gomez-Zara D, DeChurch LA, Contractor NS (2020) A taxonomy of team-assembly systems: Understanding how people use technologies to form teams. Proc ACM Hum-Comput Interaction 4(CSCW2):1–36

23. Graffelman J, Moreno V (2013) The mid p-value in exact tests for hardy-weinberg equilibrium. Stat Appl Genet Mol Biol 12(4):433–448

24. Hammer AL (1996) Mbti applications: A decade of research on the myers-briggs type indicator. Consulting Psychologists Press

25. Hwang G, Yang M (2001) An optimality theory for mid p-values in 2x2 contingency tables. Stat Sin 807-826

26. John OP, Donahue EM, Kentle RL (1991) Big five inventory. Journal of Personality and Social Psychology

27. Jones G, Chace BC, Wright J (2020) Cultural diversity drives innovation: empowering teams for success. Int J Innov Sci

28. Juárez J, Santos C, Brizuela CA (2021) A comprehensive review and a taxonomy proposal of team formation problems. ACM Comput Entertain (CSUR) 54(7):1–33

29. Kerr NL, Tindale RS (2004) Group performance and decision making. Annu Rev Psychol 55:623–655

30. Koh E, Hong H, Tan JP-L (2018) Formatively assessing teamwork in technologyenabled twenty-first century classrooms: exploratory findings of a teamwork awareness programme in singapore. Asia Pac J Educ 38(1):129–144

31. Komlenov Ž, Budimac Z, Ivanović M, Bothe K (2013) Team formation process automatization. Educon pp. 744-750

32. Krouska A, Troussas C, Virvou M (2019) Applying genetic algorithms for student grouping in collaborative learning: A synthetic literature review. Intell Decision Technol 13(4):395–406

33. Lykourentzou I, Antoniou A, Naudet Y, Dow SP (2016) Personality matters: Balancing for personality types leads to better outcomes for crowd teams. Proceedings of the 19th acm conference on computer-supported cooperative work & social computing pp. 260-273

34. Ma H, Li J, Zhu H, Tang W, Huang Z, Tang Y (2022). Collaborative optimization of learning team formation based on multidimensional characteristics and constraints modeling: A team leader-centered approach via e-cargo. IEEE Transactions on Computational Social Systems

35. Maqtary N, Mohsen A, Bechkoum K (2019) Group formation techniques in computer supported collaborative learning: A systematic literature review. Technol Knowl Learn 24:169–190

36. Mazni O, Syed-Abdullah S-L, Hussin NM (2010) Analyzing personality types to predict team performance. 2010 international conference on science and social research (cssr 2010) pp. 624-628

37. Meredith BR (2010) Management teams: Why they succeed or fail. Human Resource Management International Digest

38. Minn E (2021) Odi on employee motivation and teamwork to enhance employee performance: a case study of aaa company limited in a pharmaceutical industry. J Bangokthnburi Univ 10(1)

39. Murzi HG, Chowdhury TM, Karlovšek J, Ulloa BR (2020) Working in large teams: Measuring the impact of a teamwork model to facilitate teamwork development in engineering students working in a real project. Int J Eng Educ 36(1):274–295

40. Myers IB (1962) The myers-briggs type indicator: Manual

41. Odo C, Masthoff J, Beacham N (2019) Group formation for collaborative learning: A systematic literature review. Artificial intelligence in education: 20th international conference, aied 2019, chicago, il, usa, june 25-29, 2019, proceedings, part ii 20 pp. 206-212

42. Ounnas A, Davis H, Millard D (2008) A framework for semantic group formation. 2008 eighth ieee international conference on advanced learning technologies pp. 34-38

43. Partington D, Harris H (1999) Team role balance and team performance: an empirical study. J Manag Dev

44. Pieterse V, Kourie DG, Sonnekus IP (2006) Software engineering team diversity and performance. 2006 ann. conf. of the south african institute of computer scientists and information technologists on it research in developing countries pp. 180-186

45. Ratcheva V (2009) Integrating diverse knowledge through boundary spanning processes-the case of multidisciplinary project teams. Int J Proj Manag 27(3):206–215

46. Rosen MA, DiazGranados D, Dietz AS, Benishek LE, Thompson D, Pronovost PJ, Weaver SJ (2018) Teamwork in healthcare: Key discoveries enabling safer, highquality care. Am Psychol 73(4):433

47. Rothman KJ, Greenland S, Lash TL, et al (2008) Modern epidemiology (Vol. 3). Wolters Kluwer Health

48. Sánchez OR, Ordóñez CAC, Duque MÁR, Pinto IIBS (2021) Homogeneous group formation in collaborative learning scenarios: An approach based on personality traits and genetic algorithms. IEEE Trans Learn Technol 14(4):486–499

49. Sancho-Thomas P, Fuentes-Fernández R, Fernández-Manjón B (2009) Learning teamwork skills in university programming courses. Comput Educ 53(2):517–531

50. Sanz-Martínez L, Martínez-Monés A, Bote-Lorenzo ML, Munoz-Cristóbal JA, Dimitriadis Y (2017) Automatic group formation in a mooc based on students' activity criteria. European conference on technology enhanced learning pp. 179-193

51. Smith KA (1995) Cooperative learning: Effective teamwork for engineering classrooms. Proceedings frontiers in education 1995 25th annual conference. engineering education for the 21st century Vol. 1, pp. 2b5-13

52. Soboleva EV, Karavaev NL (2020) Characteristics of the project-based teamwork in the case of developing a smart application in a digital educational environment. Eur J Contermp Educ 9(2):417–433

53. Spoelstra H, Van Rosmalen P, Houtmans T, Sloep P (2015) Team formation instruments to enhance learner interactions in open learning environments. Comput Hum Behav 45:11–20

54. Stahl G (2006) Group cognition: Computer support for building collaborative knowledge. Mit Press Cambridge, MA

55. Sullivan KM, Dean A, Soe MM (2009) On academics: Openepi: a web-based epidemiologic and statistical calculator for public health. Public health reports 124(3):471–474

56. Syväoja T (2021) Understanding personalities as part of increasing the effectiveness of teamwork: case miela designroom

57. Tripathy M (2018) Building quality teamwork to achieve excellence in business organizations. Int Res J Manag, IT Soc Sci 5(3):1–7

58. Ugarte N, Aranzabal A, Arruarte A, Larrañaga M (2022) Using the behavioural tendency of students in a team environment for team formation. 2022 ieee frontiers in education conference (fie) pp. 1-9

59. Van den Bossche P, Gijselaers WH, Segers M, Kirschner PA (2006) Social and cognitive factors driving teamwork in collaborative learning environments: Team learning beliefs and behaviors. Small Group Res 37(5):490–521

60. Varvel T, Adams SG, Pridie SJ, Ruiz Ulloa BC (2004) Team effectiveness and individual myers-briggs personality dimensions. J Manage Eng 20(4):141–146

61. William HP (2009) Integer programming. Logic and integer programming (pp. 25-70). Springer

62. Yannibelli V, Amandi A (2011a) Forming well-balanced collaborative learning teams according to the roles of their members: An

evolutionary approach. 2011 ieee 12th international symposium on computational intelligence and informatics (cinti) pp. 265-270

63. Yannibelli V, Amandi A (2011b) A memetic algorithm for collaborative learning team formation in the context of software engineering courses. I. conference on advances in new technologies, interactive interfaces, and communicability pp. 92-103

64. Yannibelli V, Amandi A (2012) A deterministic crowding evolutionary algorithm to form learning teams in a collaborative learning context. Expert Syst Appl 39(10):8584–8592

65. Yannibelli V, Amandi A (2018) Collaborative learning team formation considering team roles: An evolutionary approach based on adaptive crossover, mutation and simulated annealing. Res Comput Sci 147:61–74

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Victor Sanchez-Anguix is currently an associate professor at Universitat Politècnica de València and a member of the Applied Optimization Systems (SOA) group at Instituto de Tecnología Informática. His research area focuses on artificial intelligence and applied optimization, with a special focus on heuristic-based automated negotiation between autonomous agents. He has also worked on the application of optimization algorithms in education, with a special focus on the operational management of educational institutions and classrooms. He has participated in more than 10 research projects, both national and international, in the areas of artificial intelligence and optimization.

Juan M. Alberola is an Associate Professor at the Universitat Politècnica de València (UPV) and is affiliated with the Valencian Research Institute for Artificial Intelligence (VRAIN). In 2013, he obtained his Ph.D. in Artificial Intelligence from UPV. His areas of interest include Multiagent Systems, simulation, optimization, teamwork, adaptation, learning, sports analytics, as well as the application of intelligent technologies in educational environments. He is the author of over 45 international publications and has worked on various national and international research projects related to Artificial Intelligence.

Alberto Palomares is an associate professor in the Department of Information Systems and Computing (DSIC) at the Universitat Politécnica de Valencia (UPV) and is the Academic Director of the Master Degree in Tourism Intelligence (MuInTur) of the UPV. His experience as a researcher is related to machine learning and artificial intelligence and is member of the Valencian Institute of Artificial Intelligence (VraIn).

Maria Dolores Teruel is PhD Associated professor at the Economics and Social Science Department at the Universidad Politècnica de Valencia (UPV), Spain. Bachelor in Tourism in 1997 and Public Relations and Publicity in 2005, she obtained a Postgraduated studies in New Technologies applied to Tourism Information Management in 2001. In 2016, she got his Phd. at the Polytechnic University of Valencia in the field of Tourism and Heritage management at UNESCO-World Heritage Sites. Her expertise concerns natural and cultural heritage communication and commercialization, tourism site promotion in a sustainable way, tourist demand studies, in practice as well as in research, in association with local, regional, and national administrations around Mediterranean countries, especially in Spain. Dr. Teruel has been part of interdisciplinary teams, that have conducted research in different countries other than Spain as Jordan, Morocco, Egypt, Honduras, Argentina or Tunissia.

Authors and Affiliations

Victor Sanchez-Anguix[1] · Juan M. Alberola[2] · Elena Del Val[2] · Alberto Palomares[2] · Maria Dolores Teruel[3]

Juan M. Alberola
jalberola@dsic.upv.es

Elena Del Val
edelval@dsic.upv.es

Alberto Palomares
apalomares@dsic.upv.es

Maria Dolores Teruel
dteruel@upvnet.upv.es

[1] Instituto Tecnológico de Informática, Grupo de Sistemas de Optimización Aplicada, Universitat Politécnica de Valencia, Camí de Vera s/n, Valencia 46022, Spain

[2] Valencia Research Institute for Artificial Intelligence, Universitat Politécnica de Valencia, Camí de Vera s/n, Valencia 46022, Spain

[3] Centro de Investigación Arquitectura, Patrimonio y Gestión para el Desarrollo Sostenible, Universitat Politécnica de Valencia, Camí de Vera s/n, Valencia 46022, Spain