

Don't Leave Anyone Behind: Achieving Team Performance Through Diversity

Ewa Andrejczuk

Artificial Intelligence Research Institute (IIIA-CSIC)
Change Management Tool S.L)
Bellaterra, Catalonia, Spain
ewa@iiia.csic.es

Juan A. Rodríguez-Aguilar

Artificial Intelligence Research Institute (IIIA-CSIC)
Bellaterra, Catalonia, Spain
jar@iiia.csic.es

Carles Sierra

Artificial Intelligence Research Institute (IIIA-CSIC)
Bellaterra, Catalonia, Spain
sierra@iiia.csic.es

Carme Roig

IES Torras i Bages
L'Hospitalet, Catalonia, Spain
mroig112@xtec.cat

Yolanda Parejo-Romero

Institut Mare de Déu de la Mercè
Barcelona, Catalonia, Spain
yparejo@xtec.cat

Abstract—In education, student teams are composed aiming at completing academic tasks and co-learning. Key factors influencing team performance are individual competencies, personality and gender. In this paper, we present a computational model to compose proficient and congenial teams based on students' personalities, gender, and competencies to perform tasks of different nature. Our model, called synergistic team composition, extends Wilde's post-Jungian method, which solely employs individuals' personalities and gender. In addition to formally present the synergistic team composition problem, we develop an approximate algorithm to solve it. That is, an algorithm that partitions student groups into teams that are diverse in competencies, personality and gender. Finally, we discuss our positive empirical results on student performance.

Index Terms—team composition, personality, competencies

I. INTRODUCTION

Working in co-operative groups is one of the fundamental tools to address the diversity in the classroom. There is theoretical evidence that supports co-operative work as key in educational processes. For Vygotsky the individual's cognitive abilities are strongly determined by the group they belong to, since the group facilitates the sharing of views among learners [40]. Piaget, on the other hand, supports co-operative work since learning is meant to solve the conflict between the individual and the environment, between the previous schemata and the new information that comes from the environment [32]. Finally, Lew Barnett's [7] proposes to merge the learning of contents and the learning of social strategies so that students learn that their individual and collective success are tied.

However, not just any team promotes learning. Teams may work less effectively than initially expected due to several reasons: a bad balance of team members' capacities, incorrect team dynamics, lack of communication, or difficult social situations. Teams also might face difficulties when some team members do not contribute as much as others. In order for learning to be productive, all teams in a classroom should be heterogeneous, that is, to be representative of the diversity of the whole class and balanced in size [26]. Also, effective

education must balance performance across teams, that is, performance should be as homogeneous as possible in the classroom: *No one should be left behind*. This is because the success of teams contributes to the students' grades. Hence, students that are in teams that are perceived as weaker might feel like being at a disadvantage. Considerable work in fields such as organisational psychology and industrial psychology has focused on various factors that influence team performance [6], [28], [41], [42]. These factors include: competencies, experiences, age and gender, as well as personality. The question is then: How to obtain a team composition that would balance well these individual properties in a school, a college, or a classroom? In this paper, we focus on this problem. Specifically, we address the following common education situation: there is a *complex task that has to be solved by different teams of students of the same size* [1]. The task requires that each team has at least one student that shows a minimum level of competence for each of a given set of competencies. We have a pool of students with varying genders, personalities, and competencies' levels. The computational problem is how to partition students into teams that are balanced in size, competencies, personality, and gender. We refer to these balanced teams as *synergistic teams*.

This paper makes the following contributions. To start with, we formalise the synergistic team composition problem (STCP) as the problem of partitioning a group of students into *synergistic teams* with limited size. We consider a *synergistic team* as a team that is both *proficient* (covers the task's competence requirements whenever possible) and *congenial* (balances gender and psychological traits). We propose an approximate local algorithm called SynTeam to solve the synergistic team composition problem via a greedy technique to match individual competencies with those required by the task, to balance the size of teams, and to diversify the psychological traits of teams' members. We benchmark our team composition method with current school practice. We perform two different experiments in an education scenario with over

252 students to show the effectiveness of our approach. In the first study, the relative improvement of teams composed by SynTeam vs teams composed by a traditional method is 29.2%. In the second study, the relative improvement is 25.3%. These results supports the use of team composition methods that exploit diversity in individuals' competencies, personalities and gender.

Outline. The rest of this paper is structured as follows. Section II discusses related work. Section III gives background on personality for our model. Section IV describes the synergistic team composition model and Section V presents the synergistic team composition problem (STCP). Section VI presents our algorithm to solve the STCP. Then, Section VII presents results of our algorithm in the context of team composition in the classroom. Finally, Section VIII discusses our approach and future work.

II. RELATED WORK

There are many works that advise on how to *handcraft* heterogeneous teams with the purpose of increasing team-based learning and improving team performance, for instance, [45] or [26].

[45] offers a manual method to divide a classroom based on students' personalities and genders. In this paper, we extend the method by adding competencies and we offer an algorithm to compose teams in an automatic way.

[26] advises to begin a team composition process by simply asking questions to a group of students. These questions are used to gather information about those competencies that are important for the successful completion of a given task. Students respond to each question either orally or with a show of hands. Then, students are lined up based on the number of required competencies that they have, derived from the answers to the questions. The ties are broken randomly. Finally, students are asked to count off down the line by the total number of teams in the class. For instance, if we want to have five teams, each student is assigned one number from the loop starting from 1 and finishing on 5 (i.e. 1, 2, 3, 4, 5, 1, 2, . . .). The procedure continues until all students have a number assigned. All "ones" become team 1, all "twos" become team 2, etc.

Some authors tried to automatise the team composition process. That is, they strive to create multiple teams that are as similar as possible with regard to the average values of multiple attributes [16]–[18], [25], [37]. As opposed to our approach, none of those works imposes heterogeneity in a direct way when composing teams. They rather only study a set of fixed constraints (such as avoiding clustering particular majors, ensuring that no female or international student is isolated on a team, etc). Additionally, compared to our approach where we compose teams for particular tasks, they do not explicitly consider the notion of task when composing balanced teams. Finally, in contrast to our work, they only present computational results rather than real-life experiments that show the influence on teams' performance.

To the best of our knowledge, the only tool available on the web supporting team composition is the Comprehensive Assessment of Team Member Effectiveness (CATME; www.catme.org) that assigns students to teams based on their responses to an online survey. Instructors create student surveys by selecting the variables from a given inventory [25]. The algorithm generates a "question score" for each variable characterizing how well each team's distribution of that variable complies with the teacher's wishes. The algorithm also generates a global "compliance score" for each team that informs how well the team complies with the teacher's desire. The higher these values the better the team. Their team composition algorithm starts by randomly assigning students to teams of the size specified, calculating question scores and compliance scores. Then, it iteratively changes the teams to attempt to maximize the minimum compliance score of the set of teams. This work is similar to our approach, however there are substantial differences. On top of differences discussed in the previous paragraph, authors do not analyse their solutions' quality. They assume that the groupings produced by their algorithm are near optimal. The analysis described in [23] demonstrates that it is unlikely the CATME algorithm finds near optimal results.

Finally, to our knowledge [20] is the only computational model that considers both personality and competencies while composing teams. They study the influence of personality on different task allocation strategies (minimizing either under-competency or over-competency). However, their method shows substantial differences with our work. Firstly, they do not propose an algorithm to compose teams based on both personality and competence, they only describe a model to evaluate teams. Secondly, gender balance is not considered in their setting. And finally, they do not evaluate their algorithm with real data (only via agent-based simulation).

III. PERSONALITY

Personality determines people's behaviour, cognition and emotion. Different theorists present their own definitions of personality and different ways to measure it based on their theoretical stance.

In organisational and industrial psychology, the researchers distinguish between two approaches to team composition, that is The Individual Traits Approach and The Team Balance Approach.

The Individual Traits Approach is based on the presumption that, when it comes to predicting a team's performance, some individual personality traits matter more than others. The most popular personality tests used to explore this approach are: the Myers-Briggs Type Indicator (MBTI) [14] and the Five Factor Model (aka FFM [15] or "Big Five" [22]).

The MBTI consists of four dichotomous dimensions that are represented on a binary scale, that is: Extraversion / Introversion (EI), Sensing / Intuition (SN), Thinking / Feeling (TF), Judging / Perceiving (JP). These dimensions are designed to indicate how individuals perceive the world and make decisions [14]. Within this approach, every person falls

into one of the sixteen possible combinations of the four letter codes, one letter representing each dimension. This approach is easy to interpret by non-psychologists, though reliance on dichotomous preference scores rather than continuous scores excessively restricts the level of statistical analysis [19].

The Five Factor Model uses five broad dimensions to describe different aspects of human personality, that is: Extraversion, Agreeableness, Conscientiousness, Emotional Stability and Openness to Experience [15]. Mohammed and Angell [27] examined student project teams whose task was to improve processes based on issues identified in organisational settings. The researchers measured Agreeableness, Conscientiousness, Extraversion, Emotional Stability, and Team Orientation of all teams using the team mean and team variability of each personality trait. Interestingly, none of those traits, when considered separately, was meaningfully connected to team performance. Additionally, according to [33], FFM personality instruments fail to detect significant sex differences in personality structures. It is also argued that the Big Five dimensions are too broad and heterogeneous, and lack the needed specificity to make accurate predictions in many real-life settings [11]. Finally, to our knowledge, there are no contributions in organisational psychology literature that have a clear team composition method based on FFM.

In the Individual Attributes approach research examines attributes on a one-at-a-time basis. It also typically suggests that some individuals are simply better working in teams than others. *We believe taking the Individual Approach is counter-intuitive as some people may work well together, while others may not and it rather depends on the compatibility between team members.*

Henceforth, some researchers in organisational psychology focus on the Team Balance Approach where they try to understand which team member attributes are best in terms of the configuration that they compose. Here, the question is not whether the team's mean on a given, single variable affects team performance (as in the research outlined above). This general approach explores if team members complement each other based on the particular composition of several attributes associated with each team member. *Surprisingly, research examining the Team Balance Approach has been very limited.*

The most prominent theory in Team Balance approach is the Belbin theory that emphasises the importance of *roles* in the team composition [5]. Some limited support for the theory has been reported in studies with very small samples (e.g. 10 teams in [38]), but in many studies the Belbin roles tend not to be related to team performance [9], [31], [39].

Another theory in Team Balance approach is the Post-Jungian Personality Theory [43], [45]. Its author, Douglas J. Wilde, suggests that both a diverse personality profile of team members and a balanced gender distribution, positively influence the effectiveness of a team. Here, effectiveness is understood as the probability of goal achievement while performing problem-solving tasks. He proposes a novel questionnaire that is a modified version of the Myers-Briggs

Type Indicator (MBTI) [13], the ‘‘Step II’’ version of Quenk, Hammer and Majors [45]. The questionnaire is short, contains only 20 quick questions (compared to the 93 MBTI questions [10]). This is very convenient for both experts designing teams and individuals doing the test since completing the test takes just a few minutes (for details of the questionnaire, see [45, p.21]). In contrast to the MBTI measure, which consists of four binary dimensions, the Post-Jungian Personality Theory uses the *numerical* data collected using the questionnaire [44]. Douglass J. Wilde claims that it covers the same psychological territory as MBTI [43]. He also suggests that the numerical data obtained through an MBTI questionnaire can be used as an input for team composition.

Similarly to the MBTI, the test is based on the psychiatrist C. G. Jung's cognitive-mode personality model [24]. It has two sets of variable pairs called psychological functions: **(1) Sensing / Intuition (SN)**, **(2) Thinking / Feeling (TF)**, and two sets of psychological attitudes: **(3) Perception / Judgment (PJ)**, **(4) Extroversion / Introversion (EI)**. Psychological functions and psychological attitudes compose together a personality. Every dimension of a personality (EI, SN, TF, PJ) is obtained by five multiple choice true/false questions. This method was never thoroughly tested, however the provided results of this method are promising, since within a decade this novel approach has tripled the fraction of Stanford teams awarded national prizes by the Lincoln Foundation [43].

IV. TEAM COMPOSITION MODEL

In this section we introduce and formalise our team composition problem.

A. Basic definitions

In our work we consider that each student is characterised by the following attributes:

- 1) **A unique identifier:** It distinguishes a student from others.
- 2) **Gender:** $\{male, female\}$ stands for their gender.
- 3) **A personality:** represented by four personality traits, each one within $[-1, 1]$.
- 4) **A set of competencies:** A competence integrates knowledge, skills, personal values, and attitudes that enable a student to act correctly in a task or situation [36]. Each student is assumed to possess a set of competencies with associated competence levels. Associated levels of competence can adjust as the student learns.

Next, we formally define the notions of personality and student.

Definition 1: A *personality profile* is a vector $\langle sn, tf, ei, pj \rangle \in [-1, 1]^4$ of personality traits.

We denote by $C = \{c_1, \dots, c_m\}$ the whole set of competencies, where each element $c_i \in C$ stands for a competence.

Definition 2: A *student* is represented as a tuple $\langle id, g, \mathbf{p}, l \rangle$ such that:

- id is the student's identifier;
- $g \in \{man, woman\}$ stands for student gender;
- \mathbf{p} is a personality profile vector;

- $l : C \rightarrow [0, 1]$ is a function that assigns the quality level of the student's outcome with respect to competence c . We will refer to $l(c)$ as the *competence level* of the student for competence c . We assume that when a student does not have a competence (or we do not know about it), the level of this competence is zero.

We will note the set of students as $A = \{a_1, \dots, a_n\}$ and we will use super-indexes to refer to students' attributes. For instance, given a student $a \in A$, id^a will refer to the *id* component of student a .

Next, we move on to the definition of a team.

Definition 3 (Team): A *team* is any subset of A with at least two students. We denote by $\mathcal{K}_A = (2^A \setminus \{\emptyset\}) \setminus \{\{a_i\} | a_i \in A\}$ the set of all possible teams from students in A .

Given a team K , we note by $w(K)$ and by $m(K)$ the number of women and men in the team respectively.

Next, we define a task type and a task. We understand a task as a particular instance of a *task type*. A task type determines the competence levels required to solve the task as well as the importance of each competence with respect to the others. Additionally, task types differ in requiring different levels of proficiency. For instance, some task types may require a high level of creativity (like asking students to design a city brochure). Others may require a highly analytical team (like tasks requiring solving math equations). Formally, a task type is defined as follows.

Definition 4: A task type τ is a tuple $\langle \lambda, \{(c_i, l_i, w_i)\}_{i \in I_\tau} \rangle$ where:

- I_τ is the index set of the required competencies.
- $\lambda \in [0, 1]$ is the importance given to proficiency; the higher the value of λ , the higher the importance for the proficiency of a team.
- $c_i \in C$ is a competence required to perform the task;
- $l_i \in [0, 1]$ is the required competence level for c_i ;
- $w_i \in [0, 1]$ is the importance of competence c_i for the success in solving an instance of task type τ ; and $\sum_{i \in I_\tau} w_i = 1$.

A task is an instance of a task type that specifies how many students must be included in a team. We define a task as follows:

Definition 5: A *task* t is a tuple $\langle \tau, m \rangle$ such that τ is a task type and m is the required number of students, where $m \geq 2$.

Henceforth, we denote by T the set of tasks and by \mathcal{T} the set of task types. Moreover, we will note as $C_\tau = \{c_i | i \in I_\tau\}$ the set of competencies required by task type τ .

Given a team and a task, we must consider how to assign competencies to students. Students must feel both accountable and useful when working in a team. Hence, each team member must be responsible for at least one competence. This is expressed as a *competence assignment* between competencies and students:

Definition 6: Given a task type τ and a team $K \in \mathcal{K}_A$, a competence assignment is a function $\eta : K \rightarrow 2^{C_\tau}$ satisfying that $C_\tau = \bigcup_{a \in K} \eta(a)$. We note by Θ_τ^K the set of competence assignments for task type τ and team K .

We define a list of students assigned to each competence as follows.

Definition 7: Given task type τ , team K , and competence assignment η , the set $\delta(c_i) = \{a \in K | c_i \in \eta(a)\}$ stands for those students responsible of competence c_i .

In this paper we treat a competence assignment problem as an optimisation problem: to have each competence assigned to at least one student and each student assigned to at least one competence so that the total *cost* of the assignment is minimal (in terms of under- and over-proficiency). Such optimisation problem can be cast and efficiently solved as a minimum cost flow problem [2]. The network model would contain $v = |K| + |C_\tau| + 2$ nodes and $e = |K| \cdot |C_\tau| + |K| + |C_\tau|$ edges. As discussed in [30], the minimum cost flow problem can be solved in $O(e \cdot \log(v) \cdot (e + v \cdot \log(v)))$ time on a network with v nodes and e arcs.

V. THE SYNERGISTIC TEAM COMPOSITION PROBLEM

In this section we define a model to evaluate and compose teams. We start by introducing the notion of *proficiency* as the matching degree between a team and a task given an assignment. Next, we move on to discuss the notion of *congeniality* as a measure of the diversity of a team. Finally, we combine these two notions to calculate the *synergistic* value of a team.

A. Evaluating team proficiency

Given a team and a task, we want to calculate the *degree of proficiency* of the team as a whole, noted u_{prof} . Our aim is to match each competence with the student(s) whose personal competence level is closer to the task competence level requirement. With this we aim at avoiding both *under-proficient* and *over-proficient* allocations as both of those scenarios are ominous for team performance. In the first case, under-proficient students may get frustrated because they do not have enough knowledge to cope with the assigned competence requirements. In the second case, *over-proficient* students may get distracted and unmotivated because of the easiness of the job they are asked to do [8]). We define the degrees of under-proficiency and over-proficiency as the distances between the competence levels required by the task and those offered by the assignment as follows.

Definition 8 (Degree of under-proficiency): Given a task type τ , a team K , and an assignment η , we define the team's degree of under-proficiency for the task as:

$$u(\eta) = \sum_{i \in I_\tau} w_i \cdot \frac{\sum_{a \in \delta(c_i)} |\min(l^a(c_i) - l_i, 0)|}{|\delta(c_i)| + 1}$$

Definition 9 (Degree of over-proficiency): Given a task type τ , a team K , and an assignment η , the team's degree of over-proficiency for the task is defined as:

$$o(\eta) = \sum_{i \in I_\tau} w_i \cdot \frac{\sum_{a \in \delta(c_i)} \max(l^a(c_i) - l_i, 0)}{|\delta(c_i)| + 1}$$

Given a competence assignment for a team, we can determine its proficiency degree to perform the task by calculating

a weighted average of team's over-proficiency and under-proficiency. The weight may be used to penalize more the team's under-proficiency, as some tasks strictly require teams to be at least as proficient as defined in the task type.

Definition 10: Given a task type τ , a team K and an assignment η , the proficiency degree of the team to perform the task is defined as:

$$u_{prof}(K) = \max_{\eta \in \Theta_{\tau}^K} (1 - (v \cdot u(\eta) + (1 - v) \cdot o(\eta))) \quad (1)$$

where $v \in [0, 1]$ is the penalty given to the under-proficiency of team K .

Notice that the larger the value of v the higher the importance of the proficiency degree of team K , while the lower the value v , the less important its under-proficiency. Therefore, if we want to penalise teams that cannot cope with the competence requirements (i.e. they are under-competent) we need to choose a large value for v . And similarly a small v to penalise teams with members clearly over-competent. Although the exact value to choose will depend on the particular task type and the goal for team composition. If the objective is to favour *effective* teams we should penalize more their under-proficiency and thus select a significantly large value for v . Given these definitions, $u_{prof}(K)$ is correctly defined for any team, task type and competence assignment:

Proposition 1: For any task type τ , team K , and $\eta \in \Theta_{\tau}^K$, $u(\eta) + o(\eta) \in [0, 1]$ and $0 \leq u_{prof}(K) < 1$.

Proof 1: Soundness is straightforward as a student cannot be over- and under-proficient at the same time.

Function $u_{prof}(K)$ is used to measure how proficient a team is for a given competence assignment. However, the degree of proficiency alone does not guarantee that the team will succeed at performing it. Therefore, in the next subsection we present an evaluation function to measure *congeniality* within teams. Unlike our measure for proficiency, which is based on considering a particular competence assignment, our congeniality measure will solely rely on the personalities and genders of the members of a team.

B. Evaluating team congeniality

Recent studies in organisational psychology show that there is a trade-off between the creative productivity caused by “meta-cognitive conflict” and “harmony” — good feeling — on a team [12]. This conflict is generated by people having different views of the world (associated with opposing personality and gender), whereas harmony comes from agreement between people with similar personalities [45].

Based on these findings Douglas J. Wilde [43] compiled heuristics to successfully compose teams. Inspired by his work we construct cognitively diverse teams using the psychological function pairs SN and TF, the psychological attitudes PJ and EI, and gender. In order to mathematically capture those heuristics, we define a novel utility function for *congeniality*, u_{con} , based on the following preferences:

- 1) It values more teams whose members' sensing-intuition (SN) and thinking-feeling (TF) personality dimensions are as diverse as possible;

- 2) It prefers teams with at least one member with positive EI, TF and PJ dimensions, namely an extrovert, thinking and judging person (called ETJ personality);
- 3) It values more teams with at least one person with negative EI dimension, namely introvert; and
- 4) It prefers gender balance within a team.

Therefore, the higher the congeniality value of a team, the more diverse the team. Formally, this team utility function is defined as follows:

$$u_{con}(K) = u_{SNTF}(K) + u_{ETJ}(K) + u_I(K) + u_{gender}(K),$$

(1) $u_{SNTF}(K) = \sigma(K, SN) \cdot \sigma(K, TF)$ measures the diversity in a team, where $\sigma(K, SN)$ and $\sigma(K, TF)$ stand for the standard deviations over the SN and TF personality traits of the members of team K . The larger the values of $\sigma(K, SN)$ and $\sigma(K, TF)$, the larger their product, and hence the larger the personality diversity along the SN and TF dimensions within a team.

(2) $u_{ETJ}(K) = \max_{a \in K^{ETJ}} [\max(\alpha \cdot \mathbf{p}, 0), 0]$ measures the utility of counting on ETJ personalities, being $K^{ETJ} = \{a \in K \mid tf^a > 0, ei^a > 0, pj^a > 0\}$ the set of students exhibiting ETJ personality, $\alpha = (0, \alpha, \alpha, \alpha)$ is a vector, and α is the importance of counting on an extrovert, thinking, and judging student (ETJ personality).

(3) $u_I(K) = \max_{a \in K} [\max(\beta \cdot \mathbf{p}, 0), 0]$ is the utility of counting on an introvert student, where $\beta = (0, 0, -\beta, 0)$ is a vector and β is the importance of introvert students.

(4) $u_{gender}(K) = \gamma \cdot \sin(\pi \cdot g(K))$ measures the importance of gender balance, where γ is a parameter to weigh the importance of gender balance, and $g(K) = \frac{w(K)}{w(K) + m(K)}$ calculates the ratio of women in a team ($w(K)$ and $m(K)$ are functions counting the number of women and men, respectively). A team K is perfectly gender-balanced iff $w(K) = m(K)$, and hence $\sin(\pi \cdot g(K)) = 1$.

C. Evaluating synergistic teams

A team K is effective solving a task when it is both *proficient* (covers the required competencies) and *congenial* (balances gender and psychological traits so that students work well together). We obtain its *synergistic value* as a weighted, linear combination of its proficiency and congeniality values as follows:

Definition 11: Given a team K , the synergistic value of team K is defined as:

$$s(K) = \lambda \cdot u_{prof}(K) + (1 - \lambda) \cdot u_{con}(K) \quad (2)$$

where $\lambda \in [0, 1]$ is the relative importance of the proficiency of team K .

In general, the higher the value of λ , the higher the importance for the proficiency of a team. The setting of the value of λ depends on the task type. For instance, task types that are difficult and performed for the first time (no experts on that matter) require a high level of creativity and exchange of ideas, and hence, congeniality should be more important than proficiency ($\lambda < 0.5$). However, for tasks where team

members need to act fast (such as sport competitions or rescue teams) it is crucial for a team to be proficient ($\lambda > 0.5$). For creative task types that require certain levels of both proficiency and congeniality (such as creating a webpage) the value of λ should be set to 0.5 (so that congeniality and proficiency are equally important).

D. The synergistic team composition problem

Given a set of students A , our goal is to split them into teams of even size so that each team, and the whole partition of students into teams, is balanced in terms of competencies, personality and gender. We shall refer to these balanced teams as *synergistic teams*, meaning that they are both congenial and proficient.

Therefore, we can regard our team composition problem as a particular type of set partition problem. We will refer to any partition of A as a team partition. However, we are interested in a particular type of team partitions, namely those where teams are constrained by size m as follows.

Definition 12: Given a set of students A , we say that a team partition P_m of A is constrained by size m , $|A| \geq m \geq 2$, iff for every team $K \in P_m$, $m \leq |K| \leq m + 1$ holds.

As $|K|/m$ is not necessarily a natural number, we may need to allow for some flexibility in team size within a partition. This is why we introduced above the condition $m \leq |K| \leq m + 1$. In practical terms, in a partition we want to have teams of sizes differing by at most one student. Henceforth, we will focus on team partitions constrained by some size. We note by $\mathcal{P}_m(A)$ the set of all team partitions of A constrained by size m .

The question is: which partition to choose? As discussed before, having one excellent team is not enough, we want all teams to be as good as possible (i.e., we want to avoid partitions where some teams perform very well and some very badly). In other words, we want to have teams that show a homogeneous behaviour so that there are no big differences in performance. To do that, we use the synergistic value of a team K , noted as $s(K)$ (presented formally in subsection V-C), as an expectation of its performance. Second, we define the overall performance of a partition as the Bernoulli-Nash product of the individual team synergistic values, as this function gives larger values to homogeneous, i.e., “fair”, solutions [29], than other functions like e.g. the sum.

Definition 13: Given a team partition P_m , the synergistic value of P_m is

$$S(P_m) = \prod_{K \in P_m} s(K) \quad (3)$$

Given this definition, we can formally define the STCP the problem of finding the partition with the largest synergistic value.

Definition 14: Given a set of students A , the *synergistic team composition problem (STCP)* is the problem of finding a team partition constrained by size m , $P_m^* \in \mathcal{P}_m(A)$, that maximises $S(P_m)$, namely:

$$P_m^* = \arg \max_{P_m \in \mathcal{P}_m(A)} S(P_m)$$

VI. SOLVING THE STCP

In this section we detail an algorithm, *SynTeam*, which solves the synergistic team formation problem.

A. Partitioning the set of students

We denote by $n = |A|$ the number of students in A and by b the total number of teams, $b = \lfloor n/m \rfloor$. Note that depending on the cardinality of A and the desired team size, the number of students in each team may vary in size. For instance, if there are $n = 7$ students in A and we want to compose duets, we split students into two duets and one triplet. In general, we define the quantity distribution of students in teams of a partition, noted $Q : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N} \times \mathbb{N} \cup (\mathbb{N} \times \mathbb{N})^2$ as:

$$Q(n, m) = \begin{cases} \{(b, m)\} & \text{if } n \geq m \text{ and } n \bmod m = 0 \\ \{(n \bmod m, m + 1), \\ (b - (n \bmod m), m)\} & \text{if } n \geq m \text{ and } n \bmod m \leq b \\ \{(0, m)\} & \text{otherwise} \end{cases} \quad (4)$$

Hence, $Q(n, m)$ is the quantity distribution of students in teams of sizes m and $m + 1$; these are called *feasible* teams. Beyond these cases, there is no way to compute a partition constrained by m (see def. 12). If so, $m' \leq m$, $m' = \lfloor n/(b + 1) \rfloor$ is the largest value smaller than m that can be used to compute partitions.

B. The SynTeam algorithm

In this section we present an approximate algorithm — *SynTeam* (see Algorithm 1). *SynTeam* quickly finds an initial partition, to subsequently improve it by performing student swaps between teams. First, it randomly orders the list of students and assigns students to teams one by one from that list following $Q(|A|, m)$ to generate an initial solution $(P, S(P), \boldsymbol{\eta})$ (line 1). The assignment of students to competencies is solved as described in subsection V-A.

Second, at each iteration, *SynTeam* generates a random neighbour of the current solution as follows (line 4). First, it randomly selects two teams from the current solution. Then, it computes the synergistic value of all partitions resulting from substituting the randomly selected teams by two new teams (and corresponding competence assignments. see Subsection V-A) formed by reordering the students of the randomly selected teams in all possible ways. It stores the best option in $(P', S(P'), \boldsymbol{\eta}')$. In addition, if the current iteration is the n_l -th—not necessarily consecutive—non-improving iteration,¹ the following more fine-grained procedure is applied to $(P, \boldsymbol{\eta})$ (line 6). In the ascending order determined by team and student indexes it tries to swap two students from two different teams. The first improving solution found this way (if any) is stored in $(P', \boldsymbol{\eta}')$ and the c_l counter, for non-consecutive non-improving iterations, is re-initialized. Finally, the algorithm stops after n_r consecutive non-improving iterations.

¹If the current solution is improved at an iteration, we refer to it as an improving iteration, a non-improving iteration otherwise.

Algorithm 1 SynTeam

Require: A ▷ The list of students
Require: n_r ▷ Max. # of consecutive non-impr. iterations
Require: n_l ▷ # of non-impr. iterations before student-swap
Ensure: (P, η) ▷ Best partition found and best assignments

```
1:  $(P, S(P), \eta) \leftarrow \text{GenerateRandomSolution}(A, Q(|A|, m))$ 
2:  $c_r \leftarrow 1, c_l \leftarrow 1$ 
3: while  $c_r \leq n_r$  do
4:    $(P', S(P'), \eta') \leftarrow \text{GenerateRandomNeighbor}(P, \eta)$ 
5:   if  $S(P') \leq S(P)$  and  $c_l = n_l$  then
6:      $(P, S(P), \eta) \leftarrow \text{ApplyImprovingSwap}(P, \eta)$ 
7:      $c_l \leftarrow 1$ 
8:   if  $S(P') > S(P)$  then
9:      $(P, S(P), \eta) \leftarrow (P', S(P'), \eta')$ 
10:     $c_r \leftarrow 1, c_l \leftarrow 1$ 
11:   else
12:      $c_r \leftarrow c_r + 1, c_l \leftarrow c_l + 1$ 
return  $(P, \eta)$ 
```

VII. EXPERIMENTAL RESULTS

In this section we discuss the experiments that we performed in order to pitch our automated team composition model (SynTeam) with the team composition performed by experts (secondary school teachers).

Below, we compare both team composition models in terms of how well they predict team performance in two different education scenarios. Since we observe that SynTeam outperforms experts at predicting team performance, we argue that it is the method of choice in the classroom.

A. Teacher Method

In current school practice in Catalonia, teachers distribute the students of a class into three sub-groups: 1) students who are capable of helping others, 2) students that are in need of help, and 3) the rest of students from the class. Each team should have at least one student from each sub-group. Only the sub-group “the rest of students” is allowed to have more students in a team. To distribute students, teachers rely on their knowledge of students, as not only good grades have to be taken into consideration, but also personality is important. For instance, a student with very good grades who lacks teamwork skills will not be included in the first group, and a disruptive student with low grades but a good disposition to work on themes that really matter to him/her and/or with a strong leadership, can instead be included in the first group. In the remaining, we refer to this method as the “Teacher Method”.

B. First Experiment - Final Group Assignment

Here we discuss the details of our first experiment.

Place of the experiment: “Institut Torras i Bages”, a state school in L’Hospitalet de Llobregat, Catalonia. It has 500 students in ages varying from 11 to 18. Collaborative work has been implemented in this school for the last 7 years with a steady and significant increase in the scores and quality of the final product that students are asked to deliver.

Time of the experiment: The experiment took five days, it was performed in June 2017.

Student and team data: The experiment was performed upon four groups of students (98 students in total): ‘1r ESO

A’ (25 students), ‘1r ESO B’ (25 students), ‘1r ESO C’ (22 students) and ‘1r ESO E’ (26 students).

Task type: Students were asked to undertake the set of interdisciplinary activities (“Trellat de Síntesi”), which is an obligatory task performed at the end of each year of the secondary education curriculum in Catalonia. In detail, students were asked to create a tourist brochure of their city with all details (collect the information about the city architecture, history, cuisine, main festivals, design the logo, design the brochure, translate parts of the brochure to English). Formally, the task type $\{(c_i, l_i, w_i)\}_{i \in [1, 7]}$ had seven equally important competencies, $w_i = 1/7$, with a maximally competence level requirement, $l_i = 1$.

Team size: three students per team.

Measuring Personality: Using computers and/or mobile phones, students answered the Post-Jungian Personality questionnaire described in Section III.

Competence measure: There are eight types of human intelligences [21], each representing different ways of processing information: Naturalist, Interpersonal, Logical/Mathematical, Visual/Spatial, Body/Kinaesthetic, Musical, Intrapersonal and Linguistic. We measured students’ intelligences using a self evaluation test introduced by [35].

Students knew that the purpose of personality and competence questionnaires was to generate heterogeneous teams, understood the task, and filled in the tests as requested.

Performance evaluation: Students worked in teams and at the end of every activity presented their work in front of a panel of three teachers that assessed the content, presentation and cooperation between team members using a standardized rubric on a scale between 1 and 10.

The procedure:

1. We split each class into two halves of similar size using random sampling;

2. We partitioned one of the halves into triplets by the expert method (15 teams in total); The other half was divided by SynTeam with $\lambda = 0.8$ as learnt in the experiment described in [4] (16 teams in total). In [4], we presented only our very first experiment performed with the purpose of validating the model and finding the best λ value for creative tasks in education. [4] did not present neither SynTeam algorithm nor the experimental results described here.

3. All teams performed “Trellat de Síntesi” and we collected the final marks of students. We calculated the arithmetic averages of team members’ marks to obtain team performances.

The results: We compared the marks obtained by students. Specifically, we calculated the geometrical average of marks for the teams in each partition. We used a geometric average to penalise more the partitions that are imbalanced (i.e. the variance in team performance is high). The teams composed by SynTeam obtained 8.1 in the scale between [1, 10], while teams composed by the expert method achieved only 7.3. The relative improvement measured by the difference between two geometric averages and divided by the possible improvement is equal to 29.2%. Hence, we observe that teams composed

by SynTeam perform far better than the teams composed by the teacher method.

C. Second Experiment - Scratch programming Task

We performed another experiment to evaluate the effectiveness of our algorithm, when it comes to more technical areas, such as programming.

Place of the experiment: This study took place in three different schools in Catalonia, that is: “Institut Broggi”, “Institut Olorda” and “Institut Torras i Bages”.

Time of the experiment: This experiment took place between March and November 2017.

Student and team data: The experiment was performed upon five groups of students in ages between 14 and 15 (154 students in total). Specifically, “Institut Broggi” (55 students), “Institut Olorda” (24 students) and “Institut Torras i Bages” (75 students).

Task: The experiment was performed during 2-hour technology classes, where students had to create a game, a story or an animation using the Scratch programming language (<https://scratch.mit.edu/>).

Personality and Competence test: Similarly to the first experiment, we used the self-evaluation questionnaires.

Team size: We divided each classroom into teams of size two, so that students were able to work in one computer together.

Performance evaluation: Scratch teams were evaluated by an independent Scratch expert that did not know the source of the teamings. She used a standardized evaluation form that contained the following elements: program usability, movement, sounds, appearance, objects, control and sensors. Each element was evaluated following the guidelines specified by the rubric in the scale from 1 to 10.

The procedure:

1. We split each class into two halves of similar size using random sampling;

2. We partitioned one of the halves into duets by SynTeam with $\lambda = 0.8$ (38 teams in total). The other half was divided by the expert method (37 teams in total);

3. All teams performed the task and we collected the final marks of students.

The results: We calculated a geometric average obtained by all teams within each method, which is equal to 5.87 for teams composed using SynTeam and 4.47 for teams composed by the expert method. The relative improvement measured by the difference between two geometric averages and divided by the possible improvement is equal to 25.3%. The observed result is statistically significant (p -value= 0.04). Hence, we observe that again teams composed by SynTeam achieved better performance than the teams composed by teachers.

VIII. CONCLUSIONS AND FUTURE WORK

In this paper we introduced a model to evaluate teams based on individuals’ competencies, genders and personalities. We proposed SynTeam, a greedy algorithm for partitioning groups of humans into proficient, gender, psychologically and

size balanced teams, which yields a good, but not necessarily optimum solution. We performed two different experiments in an education scenario with a total of 252 students. Our results show that the teams composed by SynTeam perform better than teams composed by a tutor that knows the students — their background, competencies, social and cognitive capabilities. We were not able to benchmark our algorithm against random groupings as composing teams expecting to perform worse raises an ethical question about fairness of students evaluation (final marks might be worse) and may cause drop in students’ engagement. We have implemented a freely available web-based application to solve the STCP that automatically selects which algorithm to use depending on the size of the problem. It is available here: <https://eduteams.iiia.csic.es/>.

To our knowledge, SynTeam is the first computational model to build synergistic teams based on size, competencies, personality diversity and gender balance.

In [3], we performed an empirical analysis of the quality of the results of this algorithm. The results in [3] show that for the size of the problems studied in this paper the performance is nearly optimal (beyond 98%).

An important aspect to take into consideration is the quality of the SynTeam groups as perceived by teachers, and the satisfaction of students grouped with SynTeam. Although teachers were surprised by some of the resulting groupings, they later expressed a high degree of satisfaction with the outcome of all the groups including the unexpected ones. The level of cooperative teamwork and the smoothness of the work flow in groups was overall very high. Students manifested a high level of satisfaction with the grouping, too, and said they found it easy to work with the people in their groups. It is important to note that this positive level of satisfaction was unusual from the teachers’ experience from previous years as students usually complain about teachers’ groupings. We did not measure the level of satisfaction via a survey because we did not have data from previous years to compare with, so this perception by the teachers has to be considered as a non-validated qualitative assessment of the process. We plan to thoroughly study the satisfaction level of students in future, longer, experiments with the grouping method.

Regarding future work, there is the need for considering richer and more sophisticated models to capture the various factors that influence the team composition process in the real world. We will consider how our problem relates to the constrained coalition formation framework [34]. This may help add constraints and preferences coming from experts that cannot be established by any algorithm (e.g. Anna cannot be in the same team with Josep as they used to have a romantic relationship).

ACKNOWLEDGEMENTS

This work was supported by projects CIMBVAL (MINECO, TIN2017-89758-R), AppPhil (RecerCaixa 2017), and Collectiveware (MINECO/FEDER, TIN2015-66863-C2-1-R). Andrejczuk thanks an Industrial PhD scholarship from the Generalitat de Catalunya (DI-060).

REFERENCES

- [1] Silvia T Acuña, Marta Gómez, and Natalia Juristo. How do personality, team processes and task characteristics relate to job satisfaction and software quality? *Information and Software Technology*, 51(3):627–639, 2009.
- [2] R.K. Ahuja, T.L. Magnanti, and J.B. Orlin. Network flows: theory, algorithms, and applications. 1993.
- [3] E. Andrejczuk, F Bistaffa, J. A. Rodríguez-Aguilar, C. Sierra, and C. Blum. Solving the synergistic team composition problem. In *Proceedings of the 17th Conference on Autonomous Agents and MultiAgent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, IN PRESS.
- [4] E. Andrejczuk, J. A. Rodríguez-Aguilar, C. Roig, and C. Sierra. Synergistic team composition. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, pages 1463–1465. International Foundation for Autonomous Agents and Multiagent Systems, 2017.
- [5] A. Aritzeta, S. Swailes, and B. Senior. Belbin’s team role model: Development, validity and applications for team building. *Journal of Management Studies*, 44(1):96–118, 2007.
- [6] J. Arnold and R. Randall. *Work psychology*. Pearson Education Limited., Harlow, England, 2010.
- [7] Lew Barnett. *Motivación, tratamiento de la diversidad y rendimiento académico*. Graó, Barcelona, 2003.
- [8] RM Bashur, A Hernandez, and JM Peiro. The impact of underemployment on individual and organizational performance. *Underemployment: Psychological, Economic, and Social Challenges içinde (ss. 187-213)* New York: Springer, 2011.
- [9] R. Batenburg, W. van Walbeek, and W. in der Maur. Belbin role diversity and team performance: is there a relationship? *Journal of Management Development*, 32(8):901–913, 2013.
- [10] G. J. Boyle. Myers-briggs type indicator (mbti): Some psychometric limitations. *Australian Psychologist*, 30(1):71–74, 1995.
- [11] Gregory J. Boyle. Critique of the five-factor model of personality. *Humanities and Social Sciences papers*, (Paper 297), 2008.
- [12] J. H. Bradley and F. J. Hebert. The effect of personality type on team performance. *Journal of Management Development*, 16:337–353, 1997.
- [13] I. Briggs and P.B. Myers. *Gifts Differing: Understanding Personality Type*. Mountain View, CA: Davies-Black Publishing, 1995 [1980].
- [14] I Briggs Myers, M. H. McCaulley, N. L. Quenk, and A. L. Hammer. Mbt manual (a guide to the development and use of the myers briggs type indicator), 1998.
- [15] Jr. Costa, P.T. and R.R. McCrae. *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) manual*. Odessa, FL: Psychological Assessment Resources, 1992.
- [16] Rex Cutshall, Srinagesh Gavirneni, and Kenneth Schultz. Indiana university’s kelley school of business uses integer programming to form equitable, cohesive student teams. *Interfaces*, 37(3):265–276, 2007.
- [17] Roger G Dear and Zvi Drezner. Applying combinatorial optimization metaheuristics to the golf scramble problem. *International Transactions in Operational Research*, 7(4-5):331–347, 2000.
- [18] Jacques Desrosiers, Nenad Mladenović, and Daniel Villeneuve. Design of balanced mba student teams. *Journal of the Operational Research Society*, 56(1):60–66, 2005.
- [19] A.J. DeVito. Review of the myers-briggs type indicator. *Ninth Mental Measurements Yearbook*, (1):1030–1032, 1985.
- [20] M. Farhangian, M. Purvis, M. Purvis, and T. B. R. Savarimuthu. Agent-based modeling of resource allocation in software projects based on personality and skill. In *Advances in Social Computing and Multiagent Systems*, pages 130–146. Springer, 2015.
- [21] H. Gardner. The theory of multiple intelligences. *Annals of Dyslexia*, 37(1):19–35, 1987.
- [22] L. R. Goldberg. An alternative “description of personality”: the big-five factor structure. *Journal of personality and social psychology*, 59(6):1216, 1990.
- [23] Tyson R Henry. Forming productive student groups using a massively parallel brute-force algorithm. In *Proceedings of the World Congress on Engineering and Computer Science*, volume 1, 2013.
- [24] C.G. Jung. *Psychological types*. Princeton University Press, Princeton, 1971.
- [25] Richard A Layton, Misty L Loughry, Matthew W Ohland, and George D Ricco. Design and validation of a web-based system for assigning members to teams using instructor-specified criteria. *Advances in Engineering Education*, 2(1):n1, 2010.
- [26] Larry K Michaelsen, Dean X Parmelee, Kathryn K McMahon, Ruth E Levine, and Diane M Billings. Team-based learning for health professions education: A guide to using small groups for improving learning. 2008.
- [27] S. Mohammed and L. C. Angell. Personality heterogeneity in teams: Which differences make a difference for team performance? *Small group research*, 34(6):651–677, 2003.
- [28] M. K. Mount, M. R. Barrick, and G. L. Stewart. Five-factor model of personality and performance in jobs involving interpersonal interactions. *Human Performance*, (11):145–65, 1998.
- [29] J. Nash. The Bargaining Problem. *Econometrica*, 18(2):155–162, April 1950.
- [30] J. B. Orlin. A faster strongly polynomial minimum cost flow algorithm. *Operations research*, 41(2):338–350, 1993.
- [31] D. Partington and H. Harris. Team role balance and team performance: an empirical study. *Journal of Management Development*, 18(8):694–705, 1999.
- [32] Jean Piaget. *Construction of reality in the child*. Routledge & Kegan Paul, London, 1957.
- [33] A. Poropat. The relationship between attributional style, gender and the fivefactor model of personality. *Personality and Individual Differences*, (33):1185–1201, 2002.
- [34] T. Rahwan, T. P. Michalak, E. Elkind, P. Faliszewski, J. Sroka, M. Wooldridge, and N. R. Jennings. Constrained coalition formation. In Wolfram Burgard and Dan Roth, editors, *AAAI*. AAAI Press, 2011.
- [35] Lynda Rice. *Common Sense Assessment in the Classroom*. Shell Education, 2013.
- [36] RA Roe. Competences-a key towards the integration of theory and practice in work psychology. *Gedrag en Organisatie*, 15(4):203–224, 2002.
- [37] Paul A Rubin and Lihui Bai. Forming competitively balanced teams. *IIE Transactions*, 47(6):620–633, 2015.
- [38] Barbara Senior. Team roles and team performance: is there ‘really’ a link? *Journal of occupational and organizational psychology*, 70(3):241–258, 1997.
- [39] H. van de Water, K. Ahaus, and R. Rozier. Team roles, team balance and performance. *Journal of Management Development*, 27(5):499–512, 2008.
- [40] L. S. Vygotsky. *Mind in Society: the Development of Higher Psychological Processes*. Harvard University Press, Cambridge, MA, 1978.
- [41] M. A. West. *Effective Teamwork: Practical Lessons Learned from Organizational Research*. Wiley-Blackwell, West Sussex, 2012.
- [42] K. B. White. Mis project teams: An investigation of cognitive style implications. *MIS Quarterly*, (8(2)):95–101, 1984.
- [43] D. J. Wilde. *Teamology: The Construction and Organization of Effective Teams*. Springer-Verlag, London, 2009.
- [44] D. J. Wilde. *Jung’s Personality Theory Quantified*. Springer-Verlag London, 2011.
- [45] D.J. Wilde. *Post-Jungian Personality Theory for Individuals and Teams*. SYDROSE LP, 2013.