Computers and Education: Artificial Intelligence 8 (2025) 100388

Contents lists available at ScienceDirect

Computers and Education: Artificial Intelligence

journal homepage: www.sciencedirect.com/journal/computers-and-education-artificial-intelligence

Optimising team dynamics: The role of AI in enhancing challenge-based learning participation experience and outcomes

Athina Georgara [a,*,1], Marc Santolini [b,c], Olga Kokshagina [b,d], Camila Justine Jacinta Haux [f], Desmé Jacobs [g], Gloria Biwott [f], Marcela Correa [h], Carles Sierra [a], Jose Luis Fernandez-Marquez [e], Juan A. Rodriguez-Aguilar [a]

[a] Artificial Intelligence Research Institute, Spanish National Research Council (IIIA-CSIC), Bellaterra, Spain
[b] Learning Planet Institute, Research Unit Learning Transitions (UR LT, joint unit with CY Cergy Paris University), Paris, F-75004, France
[c] School of Public Policy, Georgia Institute of Technology, Atlanta, GA, USA
[d] University of Sydney, Sydney, Australia
[e] University of Geneva, Geneva, Switzerland
[f] UNICEF Eastern and Southern Africa Regional Office, Cape Town, South Africa
[g] RLabs *Reconstructed Living Lab, Cape Town, South Africa
[h] Goodwall, London, UK

ARTICLE INFO

Keywords:
Artificial intelligence
Challenge-based learning
Participation experience
Teamwork
Relational well-being

ABSTRACT

The approach of engaging students with real-world challenges to enhance collaboration and problem-solving has attracted significant interest from scholars and practitioners across diverse disciplines. Often called Challenge-Based Learning (CBL), this educational approach emphasises developing collaborative and problem-solving skills, with significant learning occurring within team settings. Prior studies highlight the influence of team composition on the efficacy of learning outcomes, pointing out that factors such as gender diversity, personality trait diversity, and a wide range of skills affect team dynamics and performance. Despite these insights, the practical organisation of these teams remains a challenge, often reliant on ad-hoc methods driven primarily by the nature of the setting at hand. Importantly, CBL is typically assessed through the final product, neglecting the impact of CBL on how the participants experience the overall process. That is, CBL is usually considered effective if the outcome is of high quality, ignoring participants' experience and participation quality. This study investigates the potential of an Artificial Intelligence team composition algorithm to improve participation quality and outcomes in collaborative CBL environments.

1. Introduction

Challenge-Based Learning (CBL) is a collaborative learning approach widely used in education that engages students in solving real-world problems. In CBL settings, students work in teams to identify challenges, conduct research, propose solutions, and implement them within their communities or in a broader scale (Gallagher & Savage, 2023). Like other collaborative learning environments, CBL emphasises interaction among team members to achieve a shared goal (Dillenbourg, 1999). The primary objective of CBL is to develop learners' problem-solving and teamwork skills by working on a challenge. For example, the FIRST[2] programme encourages students to engage in science and become technology leaders (Boyer, 2017). Similarly, the Greenpower program[3] aims to motivate students to pursue careers in STEM (science, technology, engineering and mathematics) (Hitchcock, 2017).

On the one hand, collaboration is a central aspect within CBL, as students undertake challenges via teamwork and acquire collaborative skills through this process. For teams to be effective, team composi-

* Corresponding author.
E-mail addresses: ageorg@iiia.csic.es (A. Georgara), marc.santolini@learningplanetinstitute.org (M. Santolini), olga.kokshagina@learningplanetinstitute.org (O. Kokshagina), joseLuis.Fernandez@unige.ch (J.L. Fernandez-Marquez), jar@iiia.csic.es (J.A. Rodriguez-Aguilar).
1 Present address: University of Southampton, Southampton, UK, a.georgara@soton.ac.uk.
2 www.firstinspires.org.
3 www.greenpowerusa.net.

https://doi.org/10.1016/j.caeai.2025.100388
Received 11 December 2024; Received in revised form 4 March 2025; Accepted 7 March 2025
Available online 12 March 2025

tion plays an important role. By team composition, we refer to methods for putting individuals into teams to positively influence teams' overall performance, outcomes and learning experiences (Dissanayake et al., 2015; Woolley et al., 2010; Santolini et al., 2023). Factors such as gender diversity, variations in personality traits, and a broad spectrum of skills can have a significant impact on the team dynamics and effectiveness (Wang, 2022; Dissanayake et al., 2019; Andrejczuk, 2018). For instance, Andrejczuk (2018) finds that teams with *(i)* diverse personalities, *(ii)* balanced gender, and *(iii)* complementary skills achieve better quality outcomes when carrying out school projects. Wang (2022) reports that teams with higher diversity in *(i)* members' expertise, *(ii)* winning experiences, and *(iii)* geolocation distribution are more likely to win crowdsourcing contests. Despite these insights, the practical organisation of such teams frequently faces hurdles, predominantly relying on ad-hoc methods influenced by the application domain.

On the other hand, the effectiveness of CBL is typically evaluated based on the final product or outcome produced by participants (learners). However, *participation quality* is essential as well, specifically the impact on participants' well-being. CBL may positively impact participants' experiences, regardless of the outcome of the challenge. This calls to explore how participants' experiences can be measured. Here, we argue that to assess CBL, it is not enough to focus on the outcome; it is also important to carefully analyse participants' experiences to learn the impact of CBL on people.

Against this background, we explore how teams' composition during the CBL affects the learning experience and participation quality. More precisely, our purpose is twofold: to propose a team composition mechanism for *online* CBL and empirically study its benefits in terms of participation outcomes and experience. Hence, we explore team composition mechanisms to employ in online CBL settings, and therefore, we investigate its impact on people considering participation quality. To measure participation quality, we created a user survey to study how participants experience online CBL through their relational well-being (RWB) (White, 2016). Rather than dividing 'subjective' from 'objective', Relational Well-Being leverages a context-specific, situated approach where well-being's subjective, material and relational dimensions are revealed as co-constitutive. This construct encompasses the key characteristics of well-being that emerge from asking people what it means to them, through research predominantly in the global South, a context similar to the one of this study (White, 2015). In our context, we consider several aspects of relational well-being, such as participants' social network, their experience with the team and the possible development of project skills.

Additionally, since there is a lack of evidence on how to compose teams for CBL, we build on the literature on team composition techniques to explore teams' composition in CBL projects. In particular, this work explores whether a state-of-the-art artificial intelligence (AI) team composition algorithm is valuable in positively impacting participation quality and outcomes in online CBL settings. We build on an existing AI algorithm to form teams (Georgara et al., 2023; Georgara, 2023) that has been tested in several educational scenarios to form teams to carry out tasks in the classroom. The AI algorithm considers several aspects, such as people's competencies, personalities, and gender, to assemble complementary teams based on well-founded observations from organisational psychology (Wylde, 2013).

The CBL activity that we study is an online challenge to engage young people worldwide on the Goodwall platform[4] in collaboration with the Youth Agency Market (YOMA) initiative.[5] Goodwall is a mission-driven social enterprise that assists young people globally in preparing for their future careers. Similarly, the YOMA initiative focuses on empowering young individuals by offering them opportunities for personal growth and skill development. The CBL this research focuses on involves young people from sixteen countries registered to participate in an open challenge, where they first worked in teams and later were asked to reflect on and report their experiences, providing insights into the process. Our analysis examines the participants' relational well-being by comparing test and control groups—where the test group corresponds to participants involved in teams formed with the AI algorithm, while the control group to participants involved in randomly formed teams.

Our results indicate a generally positive impact of the AI algorithm on participation quality. Specifically, regarding the dimensions considered to measure participation quality, our findings show that relational well-being is the most positively impacted, followed by social network growth. Regarding team experience and project skill development, participants in AI-formed teams reported a feeling of "safe space" within their teams and significant improvement in some project skills development. Finally, we analysed how the algorithmic design affects the teams, considering different configurations of the algorithms. Our analysis indicates the most suitable configurations depending on the application's goal (e.g., optimising outcome versus optimising participation quality).

Hence, this paper contributes to the literature on collaborative learning, specifically on Challenge-Based Learning (CBL) (Savery, 2006), by shedding light on practices for team composition to improve both the outcomes of participation (skills development) and the participation quality. Specifically, we make the following contributions to the literature:

1. We operationalised the concept of relational well-being (White, 2016) as a measure for participation quality. We build on different aspects of relational well-being (subjective, relational and material) to assess the quality of interactions, the sense of community, and the overall satisfaction of team members with their relationships and roles within the team.
2. We empirically quantify the benefits of using the AI algorithm proposed in Georgara et al. (2023) and Georgara (2023) concerning participation quality. We observe that teams formed with the AI algorithm exhibit positive effects in all four categories that measure participation quality. Notably, aspects related to relational well-being are consistently influenced by AI, showing statistical significance at $p < 0.05$. We observe that improving collaboration skills is positively associated with using the AI algorithm. Similarly, for some project skills, participants from the AI-formed teams saw significant improvement ($p < 0.05$). Moreover, AI-formed teams reported significant growth in their social networks (with $p < 0.037$ on average). Regarding team experience, participants from AI-formed teams felt that they could freely express themselves with their teams. These findings suggest that using the AI algorithm proposed in Georgara et al. (2023) and Georgara (2023) to assemble team composition in CBL settings can boost people's participation quality.
3. We study how different algorithm configurations affect participants' experience and teams' outcomes to provide guidelines on configuring the AI algorithm for team composition in practice. Our AI algorithm relies on a hyperparameter $\alpha \in [0, 1]$ balancing between competencies and diversity. In our experiments, we set $\alpha = 0.6$, giving slightly more importance to competence than personality diversity when forming teams. We show that this is the optimal setting for our context. Furthermore, we also studied alternative algorithmic designs by varying α between 0 and 1 and predicting team performance and participants' experiences (individual answers) for each hyperparameter setting. Our results indicate that α is not only a hyperparameter that regulates the trade-off between competency and diversity when composing teams and impacts final project scores, as already observed in the literature (Andrejczuk, 2018), it also impacts subjective and relational aspects. This suggests that there is no one-size-fits-all algorithmic design and that α can serve as a control parameter that can adjust the team assembly to be more exploratory and participatory (low α) or more exploitative and outcome-focused

[4] https://goodwall.io/.
[5] https://yoma.world/.

(high α). Intermediate values $0.2 < \alpha < 0.4$ hit a sweet spot between these two designs.

The paper is organised as follows. Section 2 provides background and related work. Section 3 describes the method: the team composition AI algorithm we chose for our study, while subsection 3.3 details the methodology followed in carrying out our experiments to assess CBL. Section 4 thoroughly analyses our empirical results. Finally, Section 5 discusses the results and sets paths to future research.

2. Background and related work

This section provides background and related work. First, we introduce challenge-based learning as a pedagogical approach. Second, we discuss various team composition methods derived from the AI literature. Third, we outline the research programme that underpins the study in this paper, setting the stage for our empirical investigation.

2.1. Challenge-based learning

Challenge-based learning (CBL) focuses on skills development (Savery, 2006). The CBL is an instructional learner-centred approach that empowers learners to research, integrate theory and practice, and apply knowledge and skills to develop a viable solution to a defined problem. CBL is an active learning approach in which students gain skills and knowledge through active engagement with a real-life challenge and collaborative work on creative and sustainable solutions (Portuguez Castro & Gómez Zermeño, 2020; Malmqvist et al., 2015; Martin & Bolliger, 2018; Van Den Beemt et al., 2020). Common aspects among CBL initiatives are critical thinking, problem-solving, collaborative learning, and autonomy (Binder et al., 2017). Converse to other approaches, CBL engages learners in real-life situations. The challenges in CBL can sometimes be broad so that multiple real-life problems can be related to these specific challenges and tackled in a similar way. CBL confronts learners with open, relevant problems for which there is no obvious solution (Membrillo-Hernández et al., 2019), which requires self-direction from students. CBL is highly flexible regarding duration, intensity, and integration with additional frameworks and techniques (Gallagher & Savage, 2023).

The concept of exposing students to real-life problems, requiring collaboration and the development of solutions, has been applied for many years in fields such as engineering and sustainable development (Bootsma et al., 2014). Willis et al. (2017) collected a set of examples of innovative challenge-based learning activities to illustrate how competition and collaboration can be used to complement formal learning in the classroom. Hackathons and engineering contests, also known as Challenge-Based Innovation Initiatives, can be seen as CBL (Colombari et al., 2021). Hackathons and contests have gained prominence, empowering individuals from diverse backgrounds, skill sets, and socio-economic strata to collaborate in designing, prototyping, and implementing solutions (Beck et al., 2022). For example, in the case of Fusion Point CBL,[6] students are grouped into small multidisciplinary teams (five to six people), and each team has three coaches, one from each participating school (Martin et al., 2022). In these initiatives, participants collaboratively solve problems without prior contracts or predefined expectations of who will solve which problem (Benchoufi et al., 2018; Masselot et al., 2022). They decide whether they will join and what and whether they will contribute (Kokshagina, 2021). These collaborative learning activities happen in teams, and an increasing amount of research looks into team dynamics in learning activities (Santolini et al., 2023). Yet, the practical assembly of teams remains a challenge, often reliant on ad-hoc methods driven primarily by individual topic preferences and time commitment. When it comes to CBL, there is a lack of evidence on how teams should be composed and how to guide a team's creation. Below, we build on the literature on team composition techniques to explore teams' composition for CBL.

2.2. Team composition techniques

Forming a group that collaboratively learns is one of the most challenging tasks in the computer-science learning context (Ardaiz-Villanueva et al., 2011; Srba & Bielikova, 2015; Amara et al., 2016). Team composition mainly focuses on the group development life cycle, in order to optimise the process of forming teams (Zheng & Pinkwart, 2014), and reveal the attributes that optimally affect team composition (Graf & Bekele, 2006). Optimising team composition is mostly seen as a way to enhance the effectiveness of team's dynamics to team performance. For example, Lin et al. (2010) used particle swarm optimisation (PSO) to propose an enhanced PSO (EPSO) for composing well-structured collaborative learning teams.

This section examines the problem of team composition from the AI's point of view.[7] Similarly, the AI community considers team composition problems as optimisation problems. Overall, given a population of individuals, an AI team composition algorithm searches to assemble teams of individuals so that the distribution of resulting teams optimises some objective function. Examples of objective functions are the maximisation of teams' expected performance (Andrejczuk et al., 2019; Georgara et al., 2021), the minimisation of communication costs within teams (Lappas et al., 2009), the minimisation of individuals' workload (Anagnostopoulos et al., 2010), or the maximisation of the number of tasks a team can tackle (Capezzuto et al., 2020). Since team composition algorithms handle very large search spaces to form teams, *exact* optimisation algorithms, which guarantee optimality, can only cope with small problems (Andrejczuk et al., 2019). Instead, AI research on team composition algorithms has primarily resorted to local search (Hoos & Stützle, 2018) and meta-heuristic techniques (Blum & Roli, 2003). Such techniques allow us to explore the large search spaces of team composition problems to find *good enough* solutions.

Here, we build on the work of Andrejczuk et al. (2018), Andrejczuk (2018), Andrejczuk et al. (2019), Georgara et al. (2021, 2023) and Georgara (2023) who have provided insights into supporting team composition within cooperative learning environments. Working in cooperative groups is one of the fundamental tools to address the diversity in the classroom. There is broad consensus in the literature to support cooperative work as a key in educational processes, dating back to Piaget (1954), and Vygotskij (1979).

The work of Andrejczuk et al. (2018), Andrejczuk (2018), and Andrejczuk et al. (2019) addresses the following common situation in the classroom: *there is a task that different teams of students must solve* (Acuña et al., 2009). Teachers are presented with diverse students differing in gender, personality, and intelligence levels.[8] The computational challenge lies in forming balanced teams in size, intelligence, personality, and gender. Andrejczuk (2018) and Andrejczuk et al. (2019) introduced an AI algorithm, *SynTeam*, designed to optimise team composition by aligning individual intelligences with task requirements, ensuring team size balance, and diversifying psychological traits among team members. This approach is built on Douglass Wilde's post-Jungian method (Wylde, 2013), which proposes how to *form* teams by fostering *diversity* to facilitate team-based learning and improve team performance.

The effectiveness of the SynTeam algorithm was evaluated by Andrejczuk (2018) and Andrejczuk et al. (2019) through a comparative study of high-school student teams formed by the algorithm against those manually created by teachers using conventional methods. This

[6] https://www.esade.edu/en/learning-innovation/rambla/fusion-point .

[7] In accordance with the definition of AI-system within the EU AI Act. https://www.euaiact.com/article/3.

[8] Referring to Gartner Intelligences (Carter, 2005).

study, conducted in the context of cooperative, project-based learning with over 252 students, found that teams formed by SynTeam outperformed manually formed teams by 25.3% and 29.2% in their final grades. These findings underscore the benefits of leveraging diversity in team composition, highlighting the importance of integrating competencies, personalities, and *diversity* in educational processes.

Georgara et al. (2021) and Georgara et al. (2023) advance the research on team composition algorithms by generalising and expanding upon the work of Andrejczuk et al. (2018), Andrejczuk (2018), and Andrejczuk et al. (2019). Specifically these works investigate the problem of assigning teams to carry out *different* projects motivated by two real-world cases: (1) the assignment of high-school students teams to internships through the School-Work Alternation (SWA) programme fostered by the European Commission (Georgara et al., 2023); and (2) the allocation of teams of undergraduate students to classroom projects (Georgara, 2023). Georgara develops *Edu2Com*, a novel team composition algorithm to address these challenges (Georgara et al., 2021, 2023; Georgara, 2023). Edu2Com employs a diversity-based approach similar to SynTeam (Andrejczuk, 2018; Andrejczuk et al., 2019) but also accounts for (i) the *motivation* of each individual to work on each project, and (ii) the *social relationships* between individuals with other potential team-mates. These extensions are based on observations in Motivational Psychology and Social Sciences. On the one hand, the work in Deci et al. (2017) identifies *intrinsic* motivation (one type of motivation identified by *Self-Determination Theory*; Deci & Ryan, 1985), as the motivation type that leads individuals to better job performance: self-determined people tend to perform better at their jobs. On the other hand, empirical evidence in the Social Sciences literature indicates that teams with strong social bonds tend to exhibit high team performance (e.g., Lucius & Kuhnert, 1997; Carron et al., 2002). Furthermore, Edu2Com is equipped with a mechanism for handling competencies organised in some competence framework (e.g., ESCO, SFIA)[9] to compute the *matching degree* of the competencies of a team with the competencies required to perform a project. Last, Edu2Com is the first team composition algorithm in the literature capable of providing *explanations* about the composition of teams and their project assignments (Georgara et al., 2022a, 2022b).[10]

Empirically, Georgara (2023) complements the findings of Andrejczuk (2018) and Georgara (2023) by demonstrating that teams formed with Edu2Com, which consider diversity, motivation, and social bonds, tend to achieve better academic outcomes. Those experiments involved undergraduate students (in Computer Science at the Technical University of Crete[11]) and MBA students (at EADA Business School[12] in Barcelona). Georgara (2023) employed the Edu2Com algorithm to form teams of students and match them with different available class projects. She found out that the larger a team's *congeniality* (*diversity* —in competencies, personalities, and gender— plus *motivation* to work on project assignments), the better the expected academic performance of a team. Additional experiments comparing Edu2Com's efficiency and effectiveness against manual team composition by educational experts revealed that Edu2Com not only performs tasks significantly faster but also produces equally or more effective team-project matches (Georgara et al., 2023). From a practical perspective, Edu2Com matched teams of students to internships in the industry in significantly less time than human experts. That is, while an experienced teacher required the time of a working week to analyse the profiles of 100 students and match them with the several internship programs, the Edu2Com algorithm required less than 1 hour and 45 minutes to complete the task, providing better or at least equally suitable matches than the expert.

Prior studies on team composition from the computational perspective mostly disregard the quality of participation and participants' experience and focus on the outcomes such as skills development (Andrejczuk et al., 2019).

2.3. Participation quality

Measuring team participation quality involves evaluating multiple aspects of team dynamics and individual contributions (Cooke et al., 2015). Team performance should include factors that are broader than team output. Research suggests considering both individual-level factors (e.g., skills, attitudes) and team-level factors (e.g., team experience) when assessing participation quality (Brennan et al., 2013). We argue that team performance should include outcomes that sustain team members' ability to work with each other (e.g., Hackman (2002))—participation quality. When it comes to participation quality, prior research focuses on psychological safety as an important indicator of the quality of the team environment. Psychological safety, defined as a belief that others will respond positively to questions, mistakes, or feedback-seeking, indicates that members are engaged in behaviour supportive of innovation and learning (Edmondson, 1999). Team psychological safety refers to how team members feel they can take interpersonal risks and speak their minds without being rejected or punished (Edmondson, 1999). This psychological state emerges as a result of team-building efforts and other environmental factors.

Psychological safety relates to relational well-being (White, 2017). Relational well-being (RWB) is often used to highlight the significance to well-being of the health and quality of relationships and the work people put into maintaining them (White, 2016). It encompasses key characteristics of well-being that emerge from asking people what it means to them, through research predominantly in the global South, a context that is relevant to our study (White, 2015). These characteristics emphasise the sense of well-being as something social and collective, the importance of relationships, and a stress on materiality among others. RWB integrates subjective, material and relational aspects of well-being. Applied to our context, the subjective aspect of well-being explores elements such as confidence, a sense of agency, and the feeling of having a direct impact. The material aspect probes the perspectives of future career development related to participation and the practical skills gained from the project elaboration. The relational aspect examines changes in relationships with others, increases in social capital, and "transversal skills" pertinent to managing team dynamics and social network growth.

While many instruments and approaches are available for measuring team participation quality, researchers emphasise the need for further validation and adaptation of these tools. Relational well-being can provide a comprehensive framework for understanding team participation by focusing on the quality of relationships and interactions within a team. In this work, we will build on different aspects of relational well-being to better understand participation quality.

3. Method

In Section 3.1, we outline the research context where our research took place. In Section 3.2, we describe Edu2Com, the AI algorithm in the literature that we selected to group learners into teams. Finally, Section 3.3 describes: (i) how we employed Edu2Com in practice to group learners into teams to participate in a video competition challenge; and (ii) how we collected data regarding teams' outcomes and participation quality.

3.1. Research context: understanding the impact of AI on teamwork

The research presented in this paper is conducted within the framework of the Youth Agency Market (YOMA) initiative, which is dedicated to empowering young individuals by offering them opportunities for

[9] Extending SynTeam algorithm that could solely handle Gartner Intelligences.

[10] This is a fundamental feature for *trustworthy AI* as expressed by the EU AI Act (EU, 2023).

[11] https://www.ece.tuc.gr.

[12] https://www.eada.edu.

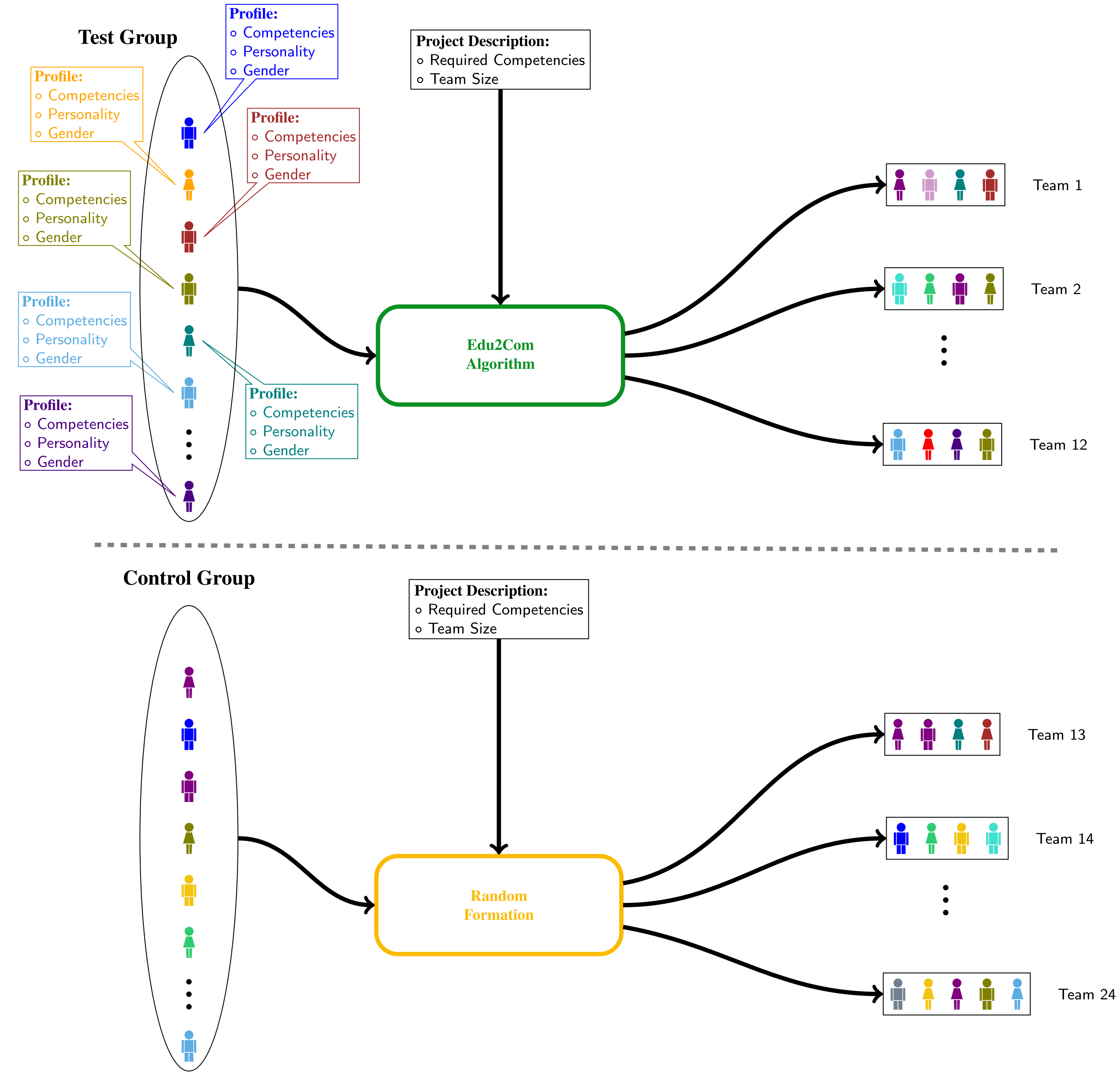

Fig. 1. Team composition workflow.

personal growth and skill development.[13] Launched in 2020, YOMA aims to mitigate the unemployment and well-being crisis among youth by providing a platform for participation in social impact tasks and pathways from learning to earning, supported by the Botnar Foundation in collaboration with Generation Unlimited (GenU) and Goodwall.[14] In 2022, The University of Geneva, in collaboration with UNICEF, led a consortium of eight partners to carry out an Operation Research project to support YOMA development. This project aims to research three major topics: (1) understanding the impact of AI on teamwork and providing learning pathways, (2) understanding the impact of using tokens for incentivising youth participation, and (3) evaluating personal and environmental impact using contextual measures of Relational Well-Being (White, 2017). The research presented in this paper addresses the use of the team composition algorithm within YOMA activities and evaluates its impact at the team and individual levels.

3.2. Edu2Com: a team composition algorithm

Although the research by Andrejczuk et al. (Andrejczuk et al., 2018; Andrejczuk, 2018; Andrejczuk et al., 2019) and Georgara et al. (Georgara et al., 2021, 2023; Georgara, 2023) have been empirically shown to be valuable for team composition in the physical classroom settings, their potential in further domains, such as online Challenge-based Learning, remains unexplored. Furthermore, and very importantly, the empirical analysis conducted by that strand of work solely focuses on the *outcome*, the teams' performance resulting from the AI-based team

[13] https://www.generationunlimited.org/yoma.
[14] https://www.goodwall.io/.

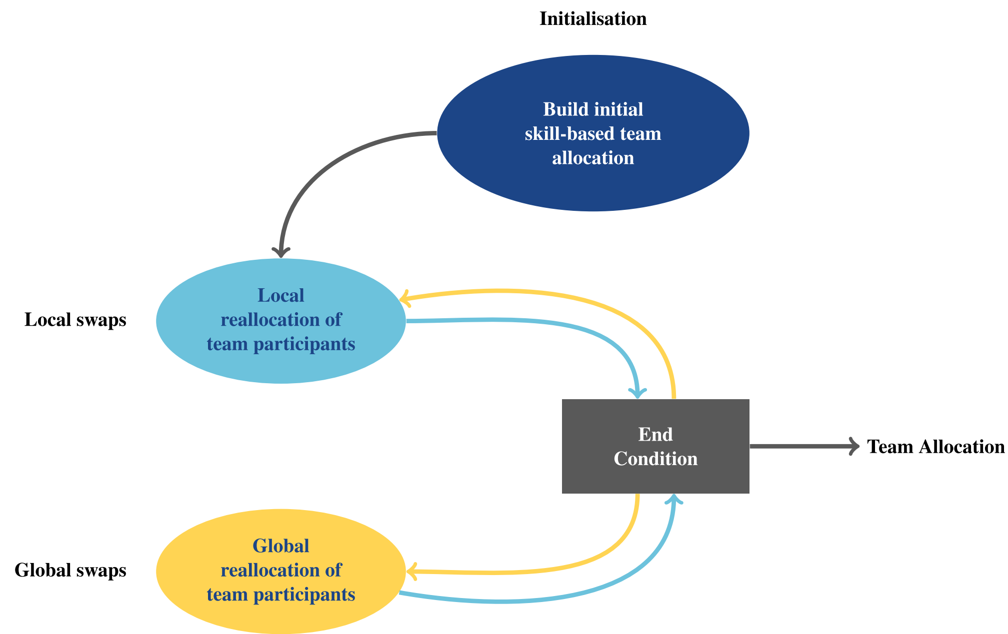

Fig. 2. Outline of the main processes the Edu2Com algorithm runs.

composition process. Therefore, the impact of those AI algorithms on people's skill development, relational well-being and overall experience with teamwork has yet to be investigated.

Given that Edu2Com, the AI team composition algorithm conceived by Georgara et al., significantly generalises the work by Andrejczuk et al. (SynTeam algorithm), in this paper, we choose Edu2Com as our team composition algorithm. We opted for Edu2Com instead of SynTeam since the former allows us to use any competence model, while the latter can handle only Gartner's Intelligences. In what follows, we outline: (1) the general workflow of Edu2Com's team composition process; and (2) the overall operation of the algorithm.

First, Fig. 1 (top) shows the team composition workflow involving Edu2Com. On the one hand, a team maker specifies a project's (or multiple projects') requirements regarding required competencies and team size. On the other hand, each person completes their *profile* by filling out tests (on personality and competencies) and reporting preferences on projects and potential teammates. Project requirements and people's profiles are inputted into the Edu2Com algorithm to output an allocation of teams to projects.

Specifically, the algorithm results in a number of teams which maximises all teams' *congeniality*. A team's congeniality comprises the team's competency to handle the project's requirements and the team's diversity in terms of personality and gender. The two elements of congeniality, the competency and the diversity, are considered in a linear combination, That is, a team's congeniality corresponds to a score where the team's competency contributes to its congeniality score by a factor α, and the diversity by a factor $1-\alpha$. Formally, the algorithm seeks a set of teams, denoted as $\mathcal{K}$, which maximises all teams' congeniality:

$$\text{congeniality}(\mathcal{K},\tau) = \prod_{K\in\mathcal{K}} (\alpha \cdot \text{competency}(K,\tau) + (1-\alpha)\cdot \text{diversity}(K)) \tag{1}$$

where τ is the project at hand, $\text{competency}(K,\tau)$ is team's $K \in \mathcal{K}$ competency for τ's requirements, $\text{diversity}(K)$ is team's K diversity in personality and gender, and α is a regulating parameter for balancing the trade-off between competency and personality. In the Appendix C, we provide further details regarding the computation of a team's competency and diversity. Notably, Section 4 provides an analysis regarding the hyperparameter α and shows the impact of α on both final project outcomes and subjective and relational aspects.

Second, Fig. 2 outlines the main processes implemented by the Edu2Com algorithm. We refer the reader to Georgara et al. (2023) and Georgara (2023) for an in-depth technical algorithm description. Edu2Com is a heuristic algorithm that matches teams of people to projects. It consists of two stages: finding an initial feasible allocation of teams to projects and iteratively improving the team allocation by swapping people between teams using different strategies. The algorithm starts by finding an efficient, feasible, and *promising* team allocation. It sequentially picks up a team for each project, from the 'hard' project to the 'simple' one. For that, it is considered a project that is 'hard' if just a few people can cover its competencies. Picking teams for the harder tasks first is a heuristic to avoid the few people that can cover it being picked by other 'simpler' projects. The second stage of Edu2Com applies several heuristics implemented as *swaps between team members*. On the one hand, Edu2Com performs *local* swaps: it randomly selects two teams to compute how to *optimally* re-distribute their team members (to maximise the *congeniality* value of both teams). On the other hand, Edu2Com considers all the teams and performs quick, systematic swaps between team members of every pair of teams looking to improve the current team allocation. Edu2Com alternates between the local and global reallocation of team members until a stopping condition occurs: either (1) no improvement occurs for several iterations; or (2) the user stops the algorithm. If so, the algorithm returns the latest team allocation. In Appendix C we provide the pseudocode of Edu2Com (see Algorithm 1). Notice that Edu2Com is an *anytime* algorithm that continuously searches for better and better team allocations rather than producing a final team allocation. The "anytime" aspect means that the user can ask the algorithm for its current best team allocation.

3.3. Using the team composition algorithm

This section outlines how the Edu2Com team composition algorithm, detailed in subsection 3.2, was applied in our context. We focus on teams of young people who work collaboratively on a creative task. The main objective was to involve these teams in conceptualising and producing a marketing video about YOMA (see subsection 3.1). To do so, we ran two *challenges* within the *Goodwall* (Bawa & Bawa, 2015) network. Goodwall is a mission-driven social enterprise that assists young people globally in preparing for their future careers. The first challenge aimed to recruit participants: the *recruitment challenge*. The second challenge involved the creation of teams of recruited participants. Each team was tasked to complete the same creative task: preparing a marketing video promoting YOMA. We will refer to this second challenge as the *competition challenge*. Next, we describe the two challenges in detail.

3.3.1. Challenges

Recruitment challenge Our first challenge was designed to recruit young individuals to participate in the competition challenge. In the initial phase, we targeted the recruitment of youth within the Goodwall network through a challenge that involved three primary activities. Firstly, participants were required to complete their Goodwall profile, detailing their educational background and any work experience they might have. Subsequently, they were prompted to share detailed information regarding their proficiency in specific skills and personality traits. Specifically, each participant was asked to complete a questionnaire comprising twenty-six questions and provide some basic personal information such as their name, age, gender, and country of origin. Following the submission of personal details, participants tackled the questionnaire. The first twenty questions were derived from the *Post-Jungian Personality test* (Wylde, 2013). The remaining six questions demanded participants to self-evaluate their proficiency across six distinct skills defined with Goodwall's skill framework (namely, the skills are *project management*, *media creation*, *using social media*, *presenting*, *storytelling* and *activism*). The questionnaire can be found in Appendix A. As soon as the participants filled out the questionnaire, they were notified with their personality test results—e.g., the scores achieved in the different personality traits dimensions. Participants were tasked with filming and sharing a brief video on the Goodwall platform discussing the personality test's results in the final part of the recruitment challenge. During the recruiting challenge, 97 people participated and completed the challenge.

Competition challenge The second challenge was a video competition in which participants worked in teams to collaboratively design and produce a marketing video showcasing and discussing the YOMA initiative. The teams competed against one another, with the winning team earning a monetary prize. The teams were then instructed to establish a communication channel to coordinate their efforts and collaboratively produce a short video. To ensure all team members had a comprehensive understanding of YOMA, each participant was required to join the YOMA platform.[15] Subsequently, each team collaboratively produced a short, two-minute video about YOMA and to post it on the Goodwall platform subsequently. These videos entered a competition, with the winning team determined by a committee that evaluated all submissions, as detailed in the methodology section on video evaluation (subsubsection 3.3.4). All 24 teams (with at least 3 team members per team) formed during the competition challenge posted a video and, therefore, participated in the competition.

3.3.2. Data description

As mentioned above, each participant provided us with personal information regarding their competency, personality, and personal details during the recruiting challenge. First, each participant indicated their competency *per skill* considering five different levels of expertise: *(1)* Novice, *(2)* Advanced Beginner, *(3)* Competent, *(4)* Proficient, and *(5)* Expert. Table 1 describes how the 97 participants assessed themselves. Notably, most participants assessed their competency with high levels of expertise. Next, each participant completed a personality test consisting of 20 questions. The personality test developed by D. Wilde aims to position an individual across 4 dimensions (personality traits): *(1)* Extroversion / Introversion, *(2)* Sensing / Intuition, *(3)* Thinking / Feeling, and *(4)* Perceiving / Judging. The test includes 5 questions per dimension, giving a score of −1 to answers indicating Extroversion, Sensing,

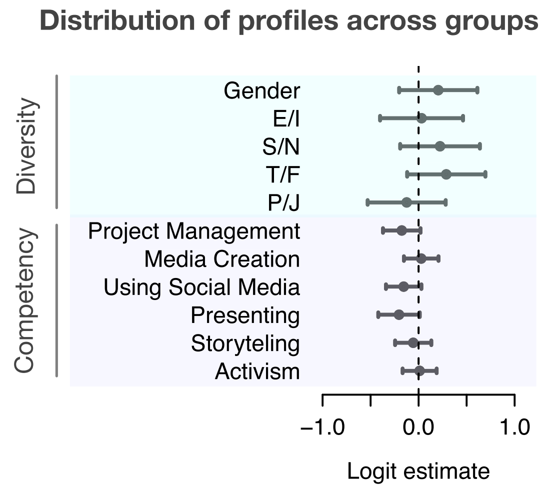

Fig. 3. Personality and gender distribution per group. We quantify the distribution of gender, personalities (Post-Jungian Personality test) and competencies between the AI and control groups using a logistic regression. We show the estimate and standard error. We find no statistical difference across characteristics, as assessed using a logistic regression between the two groups ($p > 0.1$).

Thinking and Perceiving and a score of +1 to answers indicating Introversion, Intuition, Feeling and Judging. The personality test result is a 4-dimension vector consisting of the average score for each personality trait. Negative average scores show that an individual inclines towards Extroversion / Sensing / Thinking / Perceiving, while positive average scores show that an individual leans towards Introversion / Intuition / Feeling / Judging. Regarding the participants' genders, 46 (47.4%) participants were male and 51 (52.6%) were female, ranging from 16 years old to 59 years old. Finally, the participants came from 16 different countries worldwide (including Nigeria, South Africa, India, Kenya, USA, Philippines etc.), with the majority coming from Nigeria (51.55%). Fig. 3 illustrates an overview of the data gathered.

3.3.3. Forming teams

After completing the recruiting challenge and before launching the competition challenge, we formed teams for the latter challenge. First, we split the 97 participants into two groups: the *test group* and the *control group*. The test group contained 48 participants with equal numbers of male and female participants (i.e., 24 males and 24 females). The control group contained the remaining 49 participants (22 males and 27 females). Table 2 shows the composition of the two groups in terms of gender and personality. In total, we formed 24 teams of 4 or 5 members each. Specifically, considering the test group, we formed 12 teams of 4 members following the Edu2Com algorithm described in Section 3. Considering the control group, we randomly formed 11 teams of 4 members and 1 team of 5 members.

3.3.4. Video competition: evaluating the participating teams

As mentioned in subsection 3.3, during the competition challenge, each team developed a short video and posted it on the Goodwall platform. The teams competed for the best video award, with all 24 teams successfully producing their videos within the two-week challenge period. After posting the videos on the Goodwall platform, a committee of 6 members were asked to watch and evaluate each video with a mark ranging from 5 to 10, where 5 indicated a low-quality video and 10 a high-quality one. The committee members came from the Goodwall and YOMA (R-Labs) networks. None of the committee members knew that the algorithm was used to form teams. Moreover, none of the committee members revealed their marks to the others. For the final mark, we aggregated the 6 distinct evaluations and reached a ranking of the teams based on their outcome (video quality). The winning team received a

[15] https://www.yoma.world/.

Table 1
Average competency per group.

	Control Group		Test Group	
	Average	Standard Deviation	Average	Standard Deviation
Project Management	3.583	1.127	3.776	0.985
Media Creation	3.688	1.133	3.653	1.147
Using Social Media	3.667	1.173	3.857	1.061
Presenting	3.958	1.071	4.143	0.842
Storytelling	3.813	1.104	3.878	1.073
Activism	3.688	1.223	3.673	1.088

Table 2
Composition of control and test groups.

Number of People		Test Group	Control Group
Gender	Male	24	22
	Female	24	27
Personality	Extroversion / Introversion	32 / 16	33 / 16
	Sensing / Intuition	30 / 18	28 / 21
	Thinking / Feeling	25 / 23	22 / 27
	Perceiving / Judging	23 / 25	25 / 24

prize worth 800$ and an invitation to present their promotional video in a public (virtual) ceremony held by Goodwall.

At this stage, we emphasise that the competition among teams was designed to enhance engagement with the challenge. In particular, we intentionally designed a non-challenging task (i.e., a short advertising video) for two main reasons. First, the teams were required to collaborate online, and second, they had to navigate time zone challenges, as team members could be located in any country around the world. As such, to incentivise people to participate in a relatively easy challenge, we put forward the video competition with a monetary prize. Nonetheless, our study mainly focuses on teamwork's impact on each participant instead of on the outcome, i.e., the results of the competition.

3.3.5. Post-challenge questionnaire

The primary purpose of this study is to explore the impact of teamwork on team members on an individual level. To this end, we conducted a *post-activity survey*, asking the participants to complete a questionnaire with 36 questions. The survey we designed aimed to assess participants' skill development, relational well-being, and overall teamwork experience. Specifically, it included 8 questions about each participant's experience and communication within their team, 6 questions focused on improvements in hard skills, and 9 questions evaluating the development of their teamwork-related soft skills following the questions in Strom and Strom (2011) for assessing teamwork skills for cooperative learning. The questions address four broader categories, namely: teamwork servicing, seek & share information, communication among teammates, and getting along in the team. Moreover, we considered 7 questions regarding the relational well-being of the participants, a situated construct that has been specifically designed for research in the Global South (White, 2017). The questions cover the *subjective* (e.g. self-confidence, self-worth), *relational* (e.g. coordination and social skills), and *material* (e.g. job opportunities, career development) dimensions of well-being. These questions were developed during a co-design workshop with participants from the Yoma ecosystem, following the participatory approach from (White, 2015). Finally, a set of 5 questions was included on discrimination and harassment during teamwork. We present the complete questionnaire in Appendix B.

We collected 77 individual entries to our survey, corresponding to 79.38% of the total population participating in the challenge. Moreover, we identified and corrected for 5 misspelled team names (small caps instead of capitalisation) to match team performance results. 41 individuals belonged to the control group (i.e., 83.67% of the control group population), while 36 individuals belonged to the test group (i.e., 75% of the test group population). Section 4 presents our findings from this survey. Importantly, participants completed the questionnaire before announcing the competition's winner. This measure was taken to prevent the competition outcomes from potentially biasing the responses.

3.3.6. Computation of the effect size

In this study, we wish to quantify the effect of being in the test group (AI matchmaking) compared to the control group (random assignment) across the various quantitative reports obtained through questionnaires at the individual level. The effect size (Rosenthal et al., 1994) informs us on the proportion of change observed in the value of one quantitative answer (usually a Likert scale (Likert, 1932) from 1 to 5) between the case and control groups. A large positive effect size means that this particular question elicits higher response values within the test group compared to the control group. In addition, we wish to assess the significance of this difference given the number of individuals in each group; to do so, we use confidence intervals and p-values.

We assessed the effect size using Cohen's d statistics (Cohen, 1992), obtained from the `cohen.d` function from the `effsize` library in R.[16] Cohen's d is defined as the difference between the mean values in each group divided by their pooled standard deviation:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s} \tag{2}$$

where s is defined as

$$s = \sqrt{\frac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}{n_1 + n_2 - 2}}, \tag{3}$$

n_1 and n_2 are sample sizes of each group and the variance for one of the groups is defined as

$$s_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_{1,i} - \bar{x}_1)^2, \tag{4}$$

and similarly for the other group. The effect size magnitude is assessed using the thresholds provided in Cohen (1992), with $|d| < 0.2$ being "negligible" (black colour in plots), $|d| < 0.5$ being "small" (blue colour), and $|d| < 0.8$ being "medium" (red colour). We report confidence intervals at the 95% confidence level, using non-central t-distributions. We also computed an empirical p-value by shuffling treatment assignment across individuals and computing the resulting Cohen's d_r, repeating 1,000 times. The p-value was obtained as the proportion of times that $|d_r| > |d|$. Effect sizes that lay outside of the 95% confidence interval and for which the probability to be generated at random is less than 5% (i.e. $p < 0.05$) are deemed significant. We also highlight results for which $p < 0.1$ to guide the reader on near-significant results at a milder threshold of 10%, and that might be resolved using a larger cohort.

4. Results

To answer our research question, we evaluated the algorithm's performance concerning individual participation and the quality of team

[16] https://rdocumentation.org/packages/effectsize.

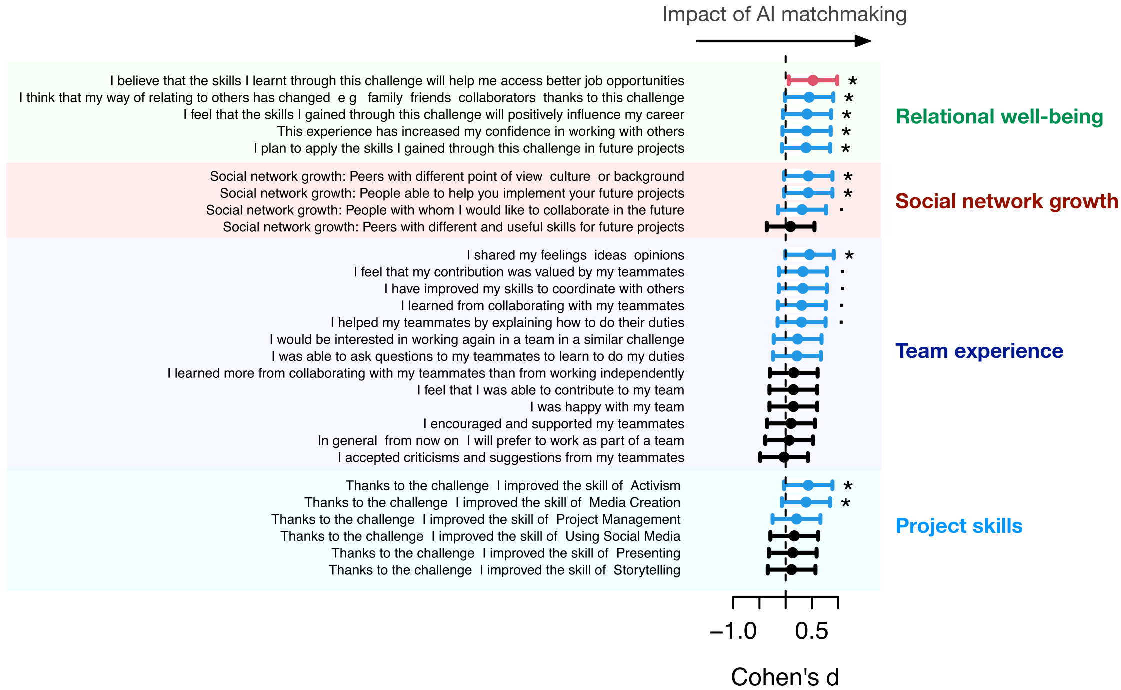

Fig. 4. Comparison of AI and control group over all the questions, grouped by category. We compute Cohen's d on the answer for each question between the AI and control groups. We show the estimate and 95% confidence interval. Significant differences are denoted by * for $p < 0.05$ and . for $p < 0.1$. The effect size magnitude is assessed using the thresholds provided in Cohen (1992), with $|d| < 0.2$ being "negligible" (black colour), $|d| < 0.5$ being "small" (blue colour), and $|d| < 0.8$ being "medium" (red colour).

performance (see Section 2.3). To assess individual participation, we conducted surveys covering various dimensions, including subjective, material, and relational aspects of well-being (Fig. 4) (White, 2017). For analytical purposes, we categorised the survey questions into four groups: relational well-being, social network growth, team dynamics, and project-related skill development. The questions are shown in Fig. 4.

To assess the impact of AI on participation quality, we computed Cohen's d effect size statistics. We show the estimates and 95% confidence intervals in Fig. 4. The effect size magnitude is assessed using the thresholds provided in Cohen (1992), with $|d| < 0.2$ being "negligible" (black colour), $|d| < 0.5$ being "small" (blue colour), and $|d| < 0.8$ being "medium" (red colour).

Several findings emerged from our analysis. Firstly, there is no evidence of a negative association, suggesting a generally positive impact of AI on the quality of participation. Aspects related to RWB are consistently influenced by AI, with all questions showing statistical significance at a level of $p < 0.05$. We find that all effect sizes are non-negligible, being either in the range $0.2 < |d| < 0.5$ (small, in blue), or $0.5 < |d| < 0.8$ (medium, red) for one question on the access to job opportunities. Participants also report significant growth in their social networks, including interactions with culturally diverse peers, which could benefit future projects. Regarding their team experience, participants felt that they could express themselves, a result that indicates a psychologically safe space to share feelings, ideas, and opinions. While barely significant, the improvement of coordination skills and the feeling of being valued in one's own contribution were also positively associated. Finally, it's noteworthy that two practical skills —–activism and media creation—– saw significant increases ($p < 0.05$). These enhancements are particularly relevant to the challenge's context, which was hosted on an impact platform promoting social good (thereby emphasising activism) and required effective use of media creation tools for success.

We further summarise the results in Fig. 5. For each category, we average across questions to provide a category-level estimate of the impact of AI. We find that RWB is overall the most impacted category, showing a medium effect size ($d = 0.55$, $p = 0.016$), followed by a close

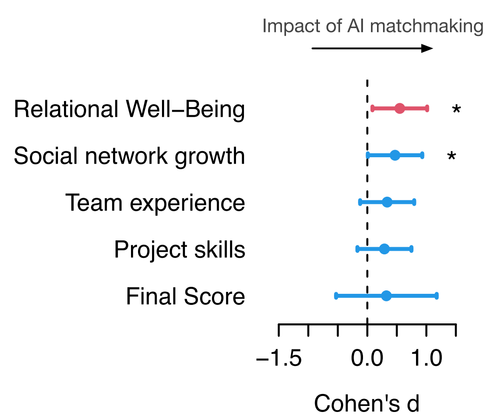

Fig. 5. Same as Fig. 4, after grouping individual answers into the 4 highlighted categories (Relational Well-Being, Social network growth, Team experience, Project skills) and showing the impact of the algorithm on team-level Final score. We use the averaged answers for each category to compute Cohen's d statistics.

to medium effect size for Social network growth ($d = 0.47$, $p = 0.037$). Small, but non-significant effects are found for the other categories. We then compare these estimates to AI's impact on the teams' final score. In this case, the statistical power is limited since we now only have 24 data points at the team level to compute associations. As such, despite a positive effect, the association of AI to the final project score is non-significant, with an effect size $d = 0.32$ more than 40% smaller than the association with relational well-being.

Finally, we investigated whether our algorithmic design was optimal. The AI algorithm relies on a balance between competency and diversity scores defined in Equation (1). This balance varies using a hyperparameter $\alpha \in [0, 1]$, controlling the proportion of competency in the total budget. This parameter was set to $\alpha = 0.6$. Nonetheless, it is possible that other designs could have been better for this context. We investigated ten alternative designs, varying the hyperparameter α between 0 and 1 by increments of 0.1. For each increment, we computed the linear

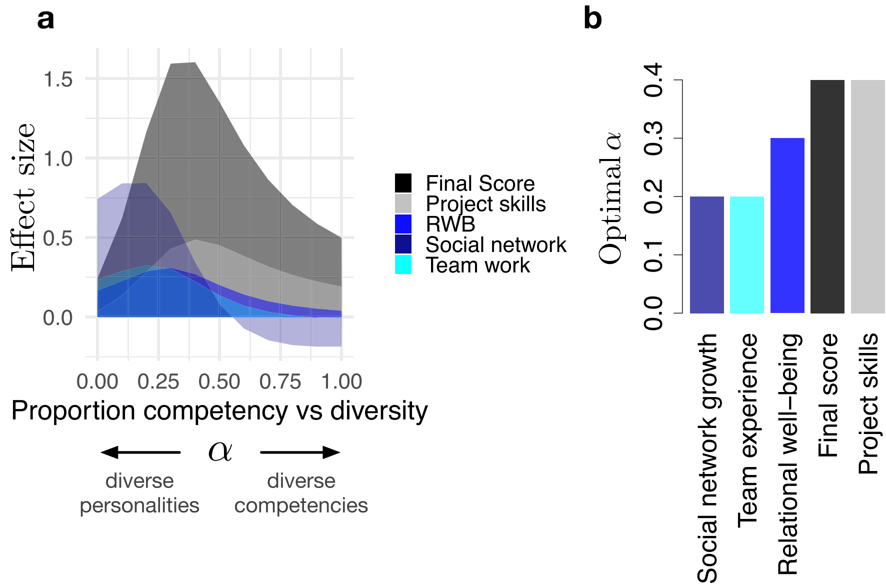

Fig. 6. Optimal design of the algorithm for various characteristics. For each team, we measured the algorithm score from Equation (1) for different proportions of competency (α) and diversity ($1-\alpha$). We then compute how these scores predict the normalised team score and the average individual answers for the 4 categories investigated. **a.** We show the effect size of a linear model predicting a given characteristic as a function of α. **b.** We show the value α with the largest effect size for each characteristic, corresponding to the optimal algorithmic design for this particular characteristic.

model $y \sim x$, where the dependent variable y is either the team performance or the average answer in each of the 4 categories of participation quality, and the independent variable x is the algorithm score ($N = 24$). The obtained estimates are shown in Fig. 6. A larger estimate means a stronger impact of the algorithm score on the dependent variable for that particular hyperparameter value. We find that $\alpha = 0.4$ is optimal for the association between the algorithmic and final project scores. Accordingly, it is also the optimal parameter when considering improving skills related to the project across individuals. In contrast, we find that the subjective and relational aspects benefit from a lower $\alpha = 0.2$. That is, individuals who belong to teams with more diverse personalities enjoy higher relational well-being, team experience and social network growth. In contrast, individuals who belong to teams with higher competency have a more successful outcome and gain more project-related skills.

5. Discussion and future work

5.1. AI and relational well-being

Our findings indicate that the application of AI for matching individuals to teams significantly impacts the quality of their experience, potentially leading to enduring benefits. In our study, participants reported a significant growth in their social capital, transversal skills, practical skills and subjective well-being. These findings indicate that team composition influences healthy dynamics at the team level.

Although we observed only a marginal and non-significant positive effect of AI on the final performance of teams, this could be attributed to the nature of the task, which may not necessitate a complex interdisciplinary integration of tasks. However, this result also indicates the limited scope of outcome-centric performance measures in the context of innovation challenges (Jaeger et al., 2023; Masselot et al., 2023). Our investigation of the subjective, relational and material aspects of well-being impacted by participation in the challenge revealed a more nuanced framework to quantify such a challenge's impact on individuals, including their future career perspective and personal development. The ability of an AI matchmaking system to yield a significant increase in the relational well-being of participants indicates that its use may be beneficial to wide-ranging settings requiring collaborative group work.

In addition, we leveraged the individual-level questionnaires on the quality of participation to investigate the impact of the hyper-parameter α of our AI algorithm (Equation (1)), controlling the balance between competency and personality diversity in the team mix, on different outcomes. We observed an interesting heterogeneity, where competency-focused matching benefits the project outcome and the acquisition of project-related skills, while diversity-focused matching benefits the team experience, social network growth and the overall team experience.

Andrejczuk (2018) also studied how to optimally set the α hyperparameter for an educational task. Andrejczuk et al.'s team composition focused on the outcome: final project scores. They empirically observed that setting α to 0.8 resulted in better predictions of team performance (nearly twice) than human experts' predictions (as thoroughly described in Section 4.5.3 in Andrejczuk (2018)). Our results indicate that α is not only a hyperparameter that regulates the trade-off between competency and diversity and impacts final project scores, it also impacts subjective and relational aspects. This suggests that there is no one-size-fits-all algorithmic design and that α can serve as a control parameter that can adjust the team assembly to be more exploratory and participatory (low α) or more exploitative and outcome-focused (high α). Intermediate values $0.2 < \alpha < 0.4$ hit a sweet spot between these two designs and are suggested for the next iterations.

5.2. Implications for educational theories and pedagogies

Collaborative group work is a fundamental approach to addressing classroom diversity. According to Vygotskij, an individual's cognitive development is heavily influenced by their social group, as it fosters the exchange of perspectives among learners (Vygotskij, 1979). Piaget views cooperative work as essential because learning arises from resolving conflicts between prior knowledge and new information from the environment (Piaget, 1954). Lastly, Sarrionandia et al. advocate for integrating content learning with social skill development, emphasising that students should understand the interdependence of their individual and collective success (Sarrionandia et al., 2003).

Although educational theories argue for the value of collaborative work, putting together students in groups is a highly intricate task. This has spurred the recent interest in leveraging AI techniques to assist teachers in composing teams of students in the classroom and in empirically investigating team composition techniques (Andrejczuk et al., 2018, 2019; Georgara et al., 2021; Georgara, 2023; Georgara et al., 2023). Although all those approaches have been empirically evaluated in several education scenarios, so far their focus has been on the outcome produced by teams, and not on the learning experience, nor on participation quality. The very same observation and criticism applies to research on CBL.

Against this background, there are multiple implications for teachers and learning designers from the results reported in this paper. First, the Edu2Com algorithm stands as a *general* tool for team composition with application to multiple education scenarios. This is the case because it has proven to be valuable in the classroom (Andrejczuk et al., 2018),[17] in school-work alternation programmes (Georgara et al., 2023), and in CBL, as shown in this paper. Second, we propose a survey to quantify team participation quality, hence going beyond an individual's skill development or a team outcome. More importantly, we provide empirical evidence to educators on the benefits of using Edu2Com concerning participation quality. We observe that teams formed with the AI algorithm exhibit positive effects in all four categories that measure participation quality. Our results provide further quantitative evidence on the value of collaborative group work in line with educational theories on the topic. Finally, we also provide guidelines to practitioners on how to configure the AI algorithm, by balancing competency and diversity, to meet various learning objectives.

5.3. Future work

As to future work, we envision two strands of research. First, we are aware that the experiments reported in this paper focus on one specific challenge. Therefore, it is worth investigating whether our results hold, in general, for different types of challenges with varying complexity. Second, we would like to evaluate the impact of using an AI algorithm for teamwork in formal education settings. Third, while there are many investigations of CBL practices and examples of CBL implementation in practice, these examples remain confined to small-scale, bottom-up case studies that often need more robust theoretical grounding (Leijon et al., 2022; Doulougeri et al., 2024) and evaluation of the effectiveness of CBL. Even if our research contributes to a better understanding of the effectiveness of team composition in CBL, further work is needed to evaluate the effectiveness of CBL approaches in different settings.

CRediT authorship contribution statement

Athina Georgara: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Data curation, Conceptualization. **Marc Santolini:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Data curation, Conceptualization. **Olga Kokshagina:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Camila Justine Jacinta Haux:** Resources, Methodology, Conceptualization. **Desmé Jacobs:** Resources, Methodology, Conceptualization. **Gloria Biwott:** Resources, Methodology, Conceptualization. **Marcela Correa:** Resources, Methodology, Conceptualization. **Carles Sierra:** Software, Methodology, Conceptualization. **Jose Luis Fernandez-Marquez:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Conceptualization. **Juan A. Rodriguez-Aguilar:** Writing – review & editing, Writing – original draft, Supervision, Software, Project administration, Methodology, Conceptualization.

Statements on open data and ethics

The study was approved by the Inserm ethical committee IRB0000-3888 under opinion number 23-995. Informed consent was obtained from all participants, and their privacy rights were strictly observed. The data can be obtained by sending request e-mails to the corresponding author.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was conducted under the YOMA OR project (OPE-02570). Athina Georgara was supported by the Generalitat de Catalunya under grant 2019DI17, and the VALAWAI project (HE-101070930). Marc Santolini and Olga Koshagina were partly supported by the French Agence Nationale de la Recherche (ANR), under grant agreement ANR-21-CE38-0002. Jose Luis Fernandez-Marquez was supported by ALBATROSS EU project funded by the European Union under G.A. Nº 101137895. Juan A. Rodriguez-Aguilar was funded by projects ACISUD (PID2022-136787NB-I00), VAE (TED2021-131295B-C31), and VALAWAI (HE-101070930).

Appendix A. Personal details, personality test & skills self-assessment

1. **Personal details**

   (1) Full Name *[Free text answer]*
   (2) Gender
   ∘ Male ∘ Female ∘ Prefer not to say
   (3) Age *[Free text answer]*
   (4) Country *[Choose one out of 195 countries]*

2. **Personality test**

   (1) You are more

	1	2	3	4	5	
Sociable	∘	∘	∘	∘	∘	Reserved

   (2) You are more

	1	2	3	4	5	
Expressible	∘	∘	∘	∘	∘	Contained

   (3) You prefer

	1	2	3	4	5	
Groups	∘	∘	∘	∘	∘	Individuals

   (4) You learn better by

	1	2	3	4	5	
Listening	∘	∘	∘	∘	∘	Reading

   (5) You are more

	1	2	3	4	5	
Talkative	∘	∘	∘	∘	∘	Quiet

   (6) You prefer more

	1	2	3	4	5	
The concrete	∘	∘	∘	∘	∘	The abstract

   (7) You prefer

	1	2	3	4	5	
Fact-finding	∘	∘	∘	∘	∘	Speculating

   (8) You are more

	1	2	3	4	5	
Practical	∘	∘	∘	∘	∘	Conceptual

   (9) You are more

	1	2	3	4	5	
Hands-on	∘	∘	∘	∘	∘	Theoretical

   (10) You prefer the

	1	2	3	4	5	
Traditional	∘	∘	∘	∘	∘	Novel

   (11) You prefer

[17] Recall that Edu2Com is a more general algorithm than the one employed in Andrejczuk et al. (2018).

1 2 3 4 5
Logic ○ ○ ○ ○ ○ Empathy

(12) You are more
1 2 3 4 5
Truthful ○ ○ ○ ○ ○ Tactful

(13) You see yourself more
1 2 3 4 5
Questioning ○ ○ ○ ○ ○ Accommodating

(14) You are more
1 2 3 4 5
Sceptical ○ ○ ○ ○ ○ Tolerant

(15) You think judges should be
1 2 3 4 5
Impartial ○ ○ ○ ○ ○ Merciful

(16) You are more
1 2 3 4 5
Casual ○ ○ ○ ○ ○ Systematic

(17) You prefer activities
1 2 3 4 5
Open-ended ○ ○ ○ ○ ○ Planned

(18) You work better
1 2 3 4 5
With pressure ○ ○ ○ ○ ○ Without pressure

(19) You prefer
1 2 3 4 5
Variety ○ ○ ○ ○ ○ Routine

(20) You are more
1 2 3 4 5
Improvisational ○ ○ ○ ○ ○ Methodical

3. **Sills assessment**

(1) How confident to you feel about skill "Project Management"
1 2 3 4 5
Novice ○ ○ ○ ○ ○ Expert

(2) How confident to you feel about skill "Media Creation"
1 2 3 4 5
Novice ○ ○ ○ ○ ○ Expert

(3) How confident to you feel about skill "Using Social Media"
1 2 3 4 5
Novice ○ ○ ○ ○ ○ Expert

(4) How confident to you feel about skill "Presenting"
1 2 3 4 5
Novice ○ ○ ○ ○ ○ Expert

(5) How confident to you feel about skill "Storytelling"
1 2 3 4 5
Novice ○ ○ ○ ○ ○ Expert

(6) How confident to you feel about skill "Activism"
1 2 3 4 5
Novice ○ ○ ○ ○ ○ Expert

Appendix B. Post-activity questionnaire

1. **Overall experience**

(1) I would be interested in working again in a team in a similar challenge
1 2 3 4 5
Strongly Disagree ○ ○ ○ ○ ○ Strongly Agree

(2) I was happy with my team
1 2 3 4 5
Strongly Disagree ○ ○ ○ ○ ○ Strongly Agree

(3) In general, from now on, I will prefer to work as part of a team
1 2 3 4 5
Strongly Disagree ○ ○ ○ ○ ○ Strongly Agree

(4) In your opinion, what is the connection between Goodwall and Yoma?
Free text answer:

2. **Communication & communication channels**

(1) What communication channels (e.g., Goodwall; instant messaging apps such as WhatsApp, or Messenger; emails, video calls, etc.) have you mostly used to work with your team?
Free text answer:

(2) Did you face any difficulties in coordinating with your team (e.g., time zone differences, incompatible personal time schedules, language barrier, etc.)
Free text answer:

3. **Hard skills**

(1) Thanks to the challenge, I improved the skill of "Project Management"
1 2 3 4 5
Not improved at all ○ ○ ○ ○ ○ Improved by far

(2) Thanks to the challenge, I improved the skill of "Media Creation"
1 2 3 4 5
Not improved at all ○ ○ ○ ○ ○ Improved by far

(3) Thanks to the challenge, I improved the skill of "Using Social Media"
1 2 3 4 5
Not improved at all ○ ○ ○ ○ ○ Improved by far

(4) Thanks to the challenge, I improved the skill of "Presenting"
1 2 3 4 5
Not improved at all ○ ○ ○ ○ ○ Improved by far

(5) Thanks to the challenge, I improved the skill of "Storytelling"
1 2 3 4 5
Not improved at all ○ ○ ○ ○ ○ Improved by far

(6) Thanks to the challenge, I improved the skill of "Activism"
1 2 3 4 5
Not improved at all ○ ○ ○ ○ ○ Improved by far

4. **Soft skills**

(1) I was able to ask questions to my teammates to learn to do my duties
1 2 3 4 5
Strongly Disagree ○ ○ ○ ○ ○ Strongly Agree

(2) I learned from collaborating with my teammates
1 2 3 4 5
Strongly Disagree ○ ○ ○ ○ ○ Strongly Agree

(3) I learned more from collaborating with my teammates than from working independently
1 2 3 4 5
Strongly Disagree ○ ○ ○ ○ ○ Strongly Agree

(4) I helped my teammates by explaining how to do their duties
1 2 3 4 5
Strongly Disagree ○ ○ ○ ○ ○ Strongly Agree

(5) I shared my feelings, ideas, opinions
1 2 3 4 5
Strongly Disagree ○ ○ ○ ○ ○ Strongly Agree

(6) I accepted criticisms and suggestions from my teammates
1 2 3 4 5
Strongly Disagree ○ ○ ○ ○ ○ Strongly Agree

(7) I encouraged and supported my teammates
1 2 3 4 5
Strongly Disagree ○ ○ ○ ○ ○ Strongly Agree

(8) Were there conflicts during the challenge
1 2 3 4 5
Strongly Disagree ○ ○ ○ ○ ○ Strongly Agree

(9) I helped to solve conflicts
1 2 3 4 5
Strongly Disagree ○ ○ ○ ○ ○ Strongly Agree

5. **Relational well-being**

(1) I feel that my contribution was valued by my teammates

	1	2	3	4	5	
Strongly Disagree	○	○	○	○	○	Strongly Agree

(2) This experience has increased my confidence in working with others

	1	2	3	4	5	
Strongly Disagree	○	○	○	○	○	Strongly Agree

(3) I think that my way of relating to others has changed (e.g., family, friends, collaborators) thanks to this challenge

	1	2	3	4	5	
Strongly Disagree	○	○	○	○	○	Strongly Agree

(4) I plan to apply the skills I gained through this challenge in future projects

	1	2	3	4	5	
Strongly Disagree	○	○	○	○	○	Strongly Agree

(5) I feel that the skills I gained through this challenge will positively influence my career

	1	2	3	4	5	
Strongly Disagree	○	○	○	○	○	Strongly Agree

(6) I believe that the skills I learnt through this challenge will help me access better job opportunities

	1	2	3	4	5	
Strongly Disagree	○	○	○	○	○	Strongly Agree

(7) I have improved my skills to coordinate with others

	1	2	3	4	5	
Strongly Disagree	○	○	○	○	○	Strongly Agree

(8) If your social network has grown, what type of connections did you gain?

- ☐ Peers with different and useful skills for future projects
- ☐ Peers with different point of view, culture, or background
- ☐ People able to help you implement your future projects
- ☐ People with whom I would like to collaborate in the future
- ☐ My social network has not grown

6. **Discrimination & harassment**

(1) I felt discriminated by my team due to my country of origin

	1	2	3	4	5	
Strongly Disagree	○	○	○	○	○	Strongly Agree

(2) I felt discriminated by my team due to my race

	1	2	3	4	5	
Strongly Disagree	○	○	○	○	○	Strongly Agree

(3) I felt discriminated by my team due to my age

	1	2	3	4	5	
Strongly Disagree	○	○	○	○	○	Strongly Agree

(4) I felt discriminated by my team due to my gender/sexual identity

	1	2	3	4	5	
Strongly Disagree	○	○	○	○	○	Strongly Agree

(5) I was sexually harassed by one or more of my teammates

○ Yes ○ No

Appendix C. Computing team's competency and diversity for Edu2Com algorithm

Team's competency

For computing a team's K competency we use the metric *competence affinity* described in Georgara et al. (2023); Georgara (2023). Let $\mathcal{A}$ be a set of agents, and each agent $a \in \mathcal{A}$ be described by $\langle C_a, l_a \rangle$ where C_a is a set of acquired competencies, and $l_a : C_a \to \mathbb{R}$ is an expertise level function. With $\tau = \langle C_\tau, l_\tau, w_\tau \rangle$ we denote a task, where C_τ is a set of required competencies, $l_\tau : C_\tau \to \mathbb{R}$ is a expertise level function indicating the least required expertise level for each competence, and $w_\tau : C_\tau \to [0,1]$ is an importance weight function indicating the importance of each competence. Moreover, we assume the existence of a competence ontology that provide us with semantic similarities between the different competencies. For our survey, we used the Goodwall's competence ontology (see Fig. 7). The similarity between two competencies is defined as:

$$sim(c,c') = e^{-\lambda \cdot l} \cdot \frac{e^{\kappa \cdot h} - e^{-\kappa \cdot h}}{e^{\kappa \cdot h} + e^{-\kappa \cdot h}}$$

Then, the competency of a team $K \subseteq \mathcal{A}$ for task τ as:

$$\text{competency}(K,\tau) = \prod_{c \in \eta^*_{\tau \to K}} \max\left\{(1 - w_\tau(c), cvg(c,a))\right\}$$

where $\eta^*_{\tau \to K} : C_\tau \to 2^K$ is the *optimal* fair competence assignment function as defined in Georgara et al. (2023); and $cvg(c,a)$ is the coverage that agent $a \in K$ can provide for competence $c \in C_\tau$ and is given as:

$$cvg(c,a) = \begin{cases} l_a(c) & \text{if } c \in C_a, \\ \max_{c' \in C_a}\left\{l_a(c') \cdot sim(c,c')\right\} & \text{otherwise.} \end{cases}$$

The optimal fair competence assignment function for a given team K working on a specific task τ is the one that maximises the overall competence coverage K achieves, while (1) each team member is responsible for covering at least one and at most χ required competence, and (2) each required competence is assigned to at least one team member. Thus, for obtaining $\eta^*_{\tau \to K}$ amounts to solving the following optimisation problem:

$$\max_{\eta_{\tau \to K}} \left(\prod_{a \in K} \prod_{c \in \eta_{\tau \to K}} cvg(c,a) \right)^{x_a^c}$$

subject to

$$\sum_{a \in K} x_a^c \geq 1 \qquad \forall\, c \in C_\tau$$

$$1 \leq \sum_{c \in C_\tau} x_a^c \leq \chi \qquad \forall\, a \in K$$

where $x_a^c \in 0,1$ is a binary decision variable indicating whether team member $a \in K$ is responsible for covering required competence $c \in C_\tau$. In the work of Georgara (2023) χ is set to be $\lceil \frac{|C_\tau|}{|K|} \rceil$ to avoid overloading some very competent team members with excessive responsibilities.

Team's diversity

To compute a team's diversity in terms of personality and gender we follow the Post-Jungian Theory (Wylde, 2013). According to this theory, in order for a team to be balanced should follow the rules below:

1. Within a team the members should be as diverse as possible regarding the sensing-intuition (SN) and thinking-feeling (TF) personality traits.
2. A team should have at least one member that leans towards the extrovert (E), the thinking (T) and the judging personality traits (i.e., being of ETJ personality).
3. A team should have at least one member that leans towards the introvert (I) personality trait.
4. A team should be balanced in gender (contain more or less equal number of females and males).

Given a set of agents $\mathcal{A}$, where each agent $a \in \mathcal{A}$ is described by a 4-value vector $\mathbf{p}_a \in [-1,1]^4$ that corresponds to a' personality traits, Andrejczuk et al. (2018) proposed a novel metric, referred to as *conge-*

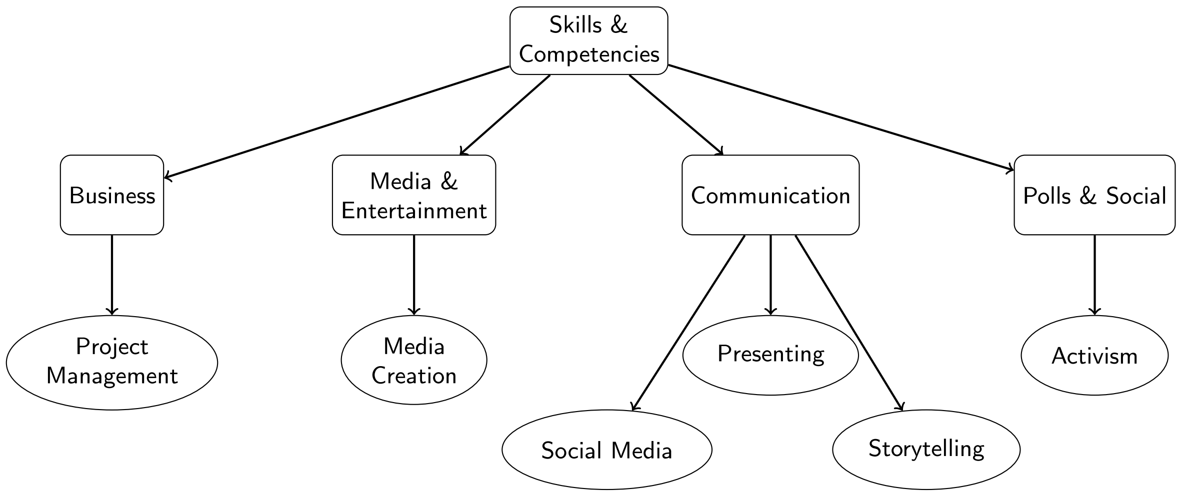

Fig. 7. Goodwall's competence ontology.

niality, that captures the above heuristics. In this study, we adopt the congeniality metric to measure the diversity of a team.

Let $K \subseteq \mathcal{A}$ be a team, for the first rule, Andrejczuk et al. use the standard deviation of the team members across the SN and the TF personality traits:

$$u_{SNTF}(K) = \sigma_{SN}(K) \cdot \sigma_{TF}(K)$$

For the second rule, the authors consider the agent $a \in K$ that belongs to the ETJ personality and contributes the most:

$$u_{ETJ}(K) = \max_{a \in K^{ETJ}} \left\{ max\{\boldsymbol{\alpha} \cdot \mathbf{p}, 0\} \right\}$$

where K^{ETJ} contains all the agents in K that belong to the ETJ personality, $\boldsymbol{\alpha} = [0, \alpha, \alpha, \alpha]$ is a vector and α is the importance of the ETJ member. Note that the personality vector encodes the personality traits as follows $\mathbf{p}_a = [SN, TF, EI, PJ]^\top$. To satisfy the third rule, the authors consider the most introvert member of the team:

$$u_I(K) = \max_{a \in K} \left\{ max\{\boldsymbol{\beta} \cdot \mathbf{p}, 0\} \right\}$$

where $\boldsymbol{\beta} = [0, 0, -\beta, 0]$ is a vector and β is the importance of the introvert (I) member. Finally, to measure gender balance, Andrejczuk et al. consider the sine of the ration of the female population within the team:

$$u_{gender}(K) = \sin\left(\frac{w(K)}{w(K) + m(K)} \cdot \pi \right)$$

where $w(K)$ is the female population in K and $m(K)$ is the male population of K (notably $w(K) + m(K) = |K|$.[18] Thus, Andrejczuk et al. (2018) defines the congeniality metric as:

$$\text{diversity}(K) = u_{SNTF}(K) + u_{ETJ}(K) + u_I(K) + \gamma \cdot u_{gender}(K)$$

where γ is a regulating parameter that indicates the importance of gender balance. In (Andrejczuk, 2018), Andrejczuk argues that $\alpha = 0.1$ (used in computing u_{ETJ}), $\beta \leq 1$ (used in computing u_I) and $\gamma = 0.1$.

We refer to Andrejczuk (2018) and Georgara (2023) for further details on competency and diversity.

[18] Georgara (2023) proposes a different metric to measure gender balance in order to make it more inclusive.

Edu2Com algorithm pseudocode

Algorithm 1: Edu2Com algorithm.

Data: Agents A, task t
Result: Task Allocation g
`/* Stage 1: Initialisation */`
$g \leftarrow \emptyset$;
sort tasks in T *wrt.* their hardness in descending order; `// Task's hardness is computed by considering the number of agents that can cover the task's required competencies`
for *task* τ **do**
  Greedily pick a team of agents $K \subseteq A$ that agents can cover best task τ;
  $g(\tau) \leftarrow K$;
  Remove agents in K from A;
end
`/* Stage 2: Iterative Improving Process */`
while *end conditions not met* **do**
  `/* Local reallocation of team participants */`
  Randomly pick two tasks τ, τ';
  $\hat{A} \leftarrow g(\tau) \cup g(\tau)$;
  `// Find the optimal teams` K^*, K'^* `for` τ `and` τ' `from` $\hat{A}$
  $K^*, K'^* \leftarrow argmax_{K,K' \subseteq \hat{A}} \left(\text{congeniality}(K, \tau) \cdot \text{congeniality}(K', \tau') \right)$;
  Update $g(\tau) \leftarrow K^*$ and $g(\tau') \leftarrow K'^*$;
  `// Global reallocation of team participants`
  for *every pair* $\tau, \tau' \in T$ **do**
    $current \leftarrow \text{congeniality}\big(g(\tau), \tau\big) \cdot \text{congeniality}\big(g(\tau'), \tau'\big)$;
    for *every pair* $a \in g(\tau)$ *and* $a' \in g(\tau')$ **do**
      `// Swap` a `with` a'
      $K \leftarrow g(\tau) \cup \{a'\} \setminus \{a\}$;
      $K' \leftarrow g(\tau') \cup \{a\} \setminus \{a'\}$;
      $new \leftarrow \text{congeniality}\big(K, \tau\big) \cdot \text{congeniality}\big(K', \tau'\big)$;
      if $new > current$ **then**
        Update $g(\tau) \leftarrow K$ and $g(\tau') \leftarrow K'$;
        Stop exhaustive pairing;
      end
    end
  end
end
return g

References

Acuña, S. T., Gómez, M., & Juristo, N. (2009). How do personality, team processes and task characteristics relate to job satisfaction and software quality? *Information and Software Technology*, *51*, 627–639.

Amara, S., Macedo, J., Bendella, F., & Santos, A. (2016). Group formation in mobile computer supported collaborative learning contexts: A systematic literature review. *Journal of Educational Technology & Society*, *19*, 258–273. https://www.jstor.org/stable/jeductechsoci.19.2.258.

Anagnostopoulos, A., Becchetti, L., Castillo, C., Gionis, A., & Leonardi, S. (2010). Power in unity: Forming teams in large-scale community systems. In *Proceedings of the 19th ACM international conference on information and knowledge management* (pp. 599–608). New York, NY, USA: Association for Computing Machinery.

Andrejczuk, E., Bistaffa, F., Blum, C., Rodríguez-Aguilar, J. A., & Sierra, C. (2019). Synergistic team composition: A computational approach to Foster diversity in teams. *Knowledge-Based Systems*, *182*, Article 104799. https://doi.org/10.1016/j.knosys.2019.06.007. https://www.sciencedirect.com/science/article/pii/S0950705119302746.

Andrejczuk, E., Rodríguez-Aguilar, J. A., Sierra, C., Roig, C., & Parejo-Romero, Y. (2018). Don't leave anyone behind: Achieving team performance through diversity. In *2018 IEEE frontiers in education conference (FIE)* (pp. 1–9). IEEE.

Andrejczuk, E. D. (2018). *Artificial intelligence methods to support people management in organisations*. PhD thesis. Barcelona, Catalunia, Spain: Universitat Autónoma de Barcelona. Retrieved from https://digital.csic.es/bitstream/10261/197543/1/Artificial_intelligence.pdf.

Ardaiz-Villanueva, O., Nicuesa-Chacón, X., Brene-Artazcoz, O., Sanz de Acedo Lizarraga, M. L., & Sanz de Acedo Baquedano, M. T. (2011). Evaluation of computer tools for idea generation and team formation in project-based learning. *Computers and Education*, *56*, 700–711. https://doi.org/10.1016/j.compedu.2010.10.012. https://www.sciencedirect.com/science/article/pii/S0360131510002976.

Bawa, O., & Bawa, T. (2015). Goodwall. https://www.goodwall.io/.

Beck, S., Bergenholtz, C., Bogers, M., Brasseur, T. M., Conradsen, M. L., Marco, D. D., Distel, A. P., Dobusch, L., Dörler, D., Effert, A., Fecher, B., Filiou, D., Frederiksen, L., Gillier, T., Grimpe, C., Gruber, M., Haeussler, C., Heigl, F., Hoisl, K., … Xu, S. M. (2022). The open innovation in science research field: A collaborative conceptualisation approach. *Industry and Innovation*, *29*, 136–185. https://doi.org/10.1080/13662716.2020.1792274.

Benchoufi, M., Fournier, M., Magrez, D., Macaux, G., Barué, V., Mansilla Sanchez, A., de Fresnoye, O., Fillaudeau, R., Tauvel-Mocquet, O., Chalabi, N., Petit-Nivard, J. F., Blondel, L., Santolini, M., & Ben Hadj Yahia, B. (2018). Epidemium: A multidisciplinary community to tackle cancer using big and open data. *Journal of Clinical Oncology*, *36*, e13604. https://doi.org/10.1200/JCO.2018.36.15_suppl.e13604.

Binder, F. V., Nichols, M., Reinehr, S., & Malucelli, A. (2017). Challenge based learning applied to mobile software development teaching. In *2017 IEEE 30th conference on software engineering education and training (CSEE&T)* (pp. 57–64).

Blum, C., & Roli, A. (2003). Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM Computing Surveys*, *35*, 268–308.

Bootsma, M. C., Vermeulen, W. J. V., Dijk, J. v., & Schot, P. P. (2014). Added value and constraints of transdisciplinary case studies in environmental science curricula. *Corporate Social-Responsibility and Environmental Management*, *21*, 155–166. https://ideas.repec.org//a/wly/corsem/v21y2014i3p155-166.html.

Boyer, N. (2017). Inspiring the next generation of scientists and engineers: K-12 and beyond. *Computer*, *50*, 17–19.

Brennan, S. E., Bosch, M., Buchan, H., & Green, S. E. (2013). Measuring team factors thought to influence the success of quality improvement in primary care: A systematic review of instruments. *Implementation Science*, *8*, 20.

Capezzuto, L., Tarapore, D., & Ramchurn, S. (2020). Anytime and efficient coalition formation with spatial and temporal constraints. In N. Bassiliades, G. Chalkiadakis, & D. de Jonge (Eds.), *Multi-agent systems and agreement technologies* (pp. 589–606). Cham: Springer International Publishing.

Carron, A. V., Colman, M. M., Wheeler, J., & Stevens, D. (2002). Cohesion and performance in sport: A meta analysis. *Journal of Sport & Exercise Psychology*, *24*, 168–188.

Carter, P. (2005). *The complete book of intelligence tests: 500 exercises to improve, upgrade and enhance your mind strength: Vol. 8*. John Wiley & Sons.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159. https://doi.org/10.1037/0033-2909.112.1.155.

Colombari, R., D'Amico, E., & Paolucci, E. (2021). Can challenge-based learning be effective online? A case study using experiential learning theory. *CERN IdeaSquare Journal of Experimental Innovation*, *5*, 40–48. https://e-publishing.cern.ch/index.php/CIJ/article/view/1287.

Cooke, N. J., Hilton, M. L., Behavioral, B., Council, N. R., et al. (2015). Overview of the research on team effectiveness. In *Enhancing the effectiveness of team science*. National Academies Press (US).

Deci, E. L., Olafsen, A. H., & Ryan, R. M. (2017). Self-determination theory in work organizations: The state of a science. *Annual Review of Organizational Psychology and Organizational Behavior*, *4*, 19–43. https://doi.org/10.1146/annurev-orgpsych-032516-113108.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York, NY: Springer.

Dillenbourg, P. (1999). What do you mean by collaborative learning? In *Collaborative-learning: Cognitive and computational approaches*. Elsevier.

Dissanayake, I., Nerur, S., & Zhang, J. (2019). Team formation and performance in online crowdsourcing competitions: The role of homophily and diversity in solver characteristics. In *ICIS 2019 proceedings*. https://aisel.aisnet.org/icis2019/crowds_social/crowds_social/5.

Dissanayake, I., Zhang, J., & Gu, B. (2015). Task division for team success in crowdsourcing contests: Resource allocation and alignment effects. *Journal of Management Information Systems*, *32*, 8–39. https://doi.org/10.1080/07421222.2015.1068604.

Doulougeri, K., Vermunt, J. D., Bombaerts, G., & Bots, M. (2024). Challenge-based learning implementation in engineering education: A systematic literature review. *Journal of Engineering Education*, *113*, 1076–1106. https://doi.org/10.1002/jee.20588. https://onlinelibrary.wiley.com/doi/abs/10.1002/jee.20588.

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, *44*, 350–383.

EU (2023). Artificial intelligence act: Deal on comprehensive rules for trustworthy AI. https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai. (Accessed 5 March 2024).

Gallagher, S. E., & Savage, T. (2023). Challenge-based learning in higher education: An exploratory literature review. *Teaching in Higher Education*, *28*, 1135–1157. https://doi.org/10.1080/13562517.2020.1863354.

Georgara, A. (2023). *Trustworthy task allocation for human teams*. PhD thesis. Barcelona, Catalunia, Spain: Universitat Autónoma de Barcelona. Retrieved from https://www.iiia.csic.es/media/publications/Trustworth_Task_Allocation_for_Human_Teams.pdf.

Georgara, A., Kazhamiakin, R., Mich, O., Palmero Aprosio, A., Pazzaglia, J. C., Rodríguez Aguilar, J. A., & Sierra, C. (2023). The AI4Citizen pilot: Pipelining AI-based technologies to support school-work alternation programmes. *Applied Intelligence*, *53*, 24157–24186.

Georgara, A., Rodríguez-Aguilar, J. A., & Sierra, C. (2021). Towards a competence-based approach to allocate teams to tasks. In *Proceedings of the 20th international conference on autonomous agents and multiagent systems* (pp. 1504–1506). Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.

Georgara, A., Rodríguez-Aguilar, J. A., & Sierra, C. (2022a). Building contrastive explanations for multi-agent team formation. In *Proceedings of the 21st international conference on autonomous agents and multiagent systems* (pp. 516–524).

Georgara, A., Rodríguez-Aguilar, J. A., & Sierra, C. (2022b). Privacy-aware explanations for team formation. In *International conference on principles and practice of multi-agent systems* (pp. 543–552). Springer.

Graf, S., & Bekele, R. (2006). Forming heterogeneous groups for intelligent collaborative learning systems with ant colony optimization. In M. Ikeda, K. D. Ashley, & T. W. Chan (Eds.), *Intelligent tutoring systems* (pp. 217–226). Berlin, Heidelberg: Springer.

Hackman, J. R. (2002). *Leading teams: Setting the stage for great performances*. Harvard Business Press.

Hitchcock, L. (2017). Greenpower: Racing to a stem finish. *Computer*, *50*, 20–22.

Hoos, H. H., & Stützle, T. (2018). Stochastic local search. In *Handbook of approximation algorithms and metaheuristics* (pp. 297–307). Chapman and Hall/CRC.

Jaeger, J., Masselot, C., Greshake Tzovaras, B., Senabre Hidalgo, E., Haklay, M., & Santolini, M. (2023). An epistemology for democratic citizen science. *Royal Society Open Science*, *10*, Article 231100. https://doi.org/10.1098/rsos.231100. https://royalsocietypublishing.org/doi/10.1098/rsos.231100.

Kokshagina, O. (2021). Open COVID19: Organizing an extreme crowdsourcing campaign to tackle grand challenges. *R & D Management*, Article radm.12470. https://doi.org/10.1111/radm.12470.

Lappas, T., Liu, K., & Terzi, E. (2009). Finding a team of experts in social networks. In *Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 467–476). New York, NY, USA: Association for Computing Machinery.

Leijon, M., Gudmundsson, P., Staaf, P., & Christersson, C. (2022). Challenge based learning in higher education – a systematic literature review. *Innovations in Education and Teaching International*, *59*, 609–618. https://doi.org/10.1080/14703297.2021.1892503.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*.

Lin, Y. T., Huang, Y. M., & Cheng, S. C. (2010). An automatic group composition system for composing collaborative learning groups using enhanced particle swarm optimization. *Computers and Education*, *55*, 1483–1493. https://doi.org/10.1016/j.compedu.2010.06.014. https://www.sciencedirect.com/science/article/pii/S0360131510001740.

Lucius, R. H., & Kuhnert, K. W. (1997). Using sociometry to predict team performance in the work place. *The Journal of Psychology*, *131*, 21–32.

Malmqvist, J., Rådberg, K. K., & Lundqvist, U. (2015). Comparative analysis of challenge-based learning experiences. In *Proceedings of the 11th international CDIO conference*. Chengdu, Sichuan, P.R. China: Chengdu University of Information Technology.

Martin, D. A., Herzog, C., Papageorgiou, K., & Bombaerts, G. (2022). Three European experiences of cocreating ethical solutions to real-world problems through challenge based learning. In *The emerald handbook of challenge based learning* (pp. 251–279). Limited: Emerald Publishing.

Martin, F., & Bolliger, D. U. (2018). Engagement matters: Student perceptions on the importance of engagement strategies in the online learning environment. *Online Learning*, *22*. https://doi.org/10.24059/olj.v22i1.1092. https://olj.onlinelearningconsortium.org/index.php/olj/article/view/1092.

Masselot, C., Jeyaram, R., Tackx, R., Fernandez-Marquez, J. L., Grey, F., & Santolini, M. (2023). Collaboration and performance of citizen science projects addressing the sustainable development goals. *Citizen Science: Theory and Practice*, *8*(1), 45. https://doi.org/10.5334/cstp.565. https://theoryandpractice.citizenscienceassociation.org/articles/10.5334/cstp.565.

Masselot, C., Tzovaras, B. G., Graham, C. L. B., Finnegan, G., Jeyaram, R., Vitali, I., Landrain, T., & Santolini, M. (2022). Implementing the co-immune open innovation program to address vaccination hesitancy and access to vaccines: Retrospective study. *Journal of Participatory Medicine*, *14*, Article e32125. https://doi.org/10.2196/32125. https://jopm.jmir.org/2022/1/e32125.

Membrillo-Hernández, J., Ramírez-Cadena, M., Martínez-Acosta, M., Cruz-Gómez, E., Muñoz-Díaz, E., & Elizalde, H. (2019). Challenge based learning: The importance of world-leading companies as training partners. *International Journal on Interactive Design and Manufacturing*, *13*, 1103–1113.

Piaget, J. (1954). *The construction of reality in the child*. New York, NY, US: Basic Books. xiii+386 pp.

Portuguez Castro, M., & Gómez Zermeño, M. G. (2020). Challenge based learning: Innovative pedagogy for sustainability through e-learning in higher education. *Sustainability*, *12*(10), 4063. https://doi.org/10.3390/su12104063. https://www.mdpi.com/2071-1050/12/10/4063.

Rosenthal, R., Cooper, H., & Hedges, L. (1994). *The handbook of research synthesis*. Russell Sage Foundation (pp. 231–244).

Santolini, M., Blondel, L., Palmer, M. J., Ward, R. N., Jeyaram, R., Brink, K. R., Krishna, A., & Barabasi, A. L. (2023). iGEM: A model system for team science and innovation. arXiv:2310.19858 [physics].

Sarrionandia, G. E., Domingo, J. R., Iniesta, F. L., Calleja, M. M., Dalmau, M. R., Maristany, C. M., del Carmen Martín, M., Pairó, N. S., Barnett, L., Pujol, N. E., et al. (2003). *Motivación, tratamiento de la diversidad y rendimiento académico: El aprendizaje cooperativo: Vol. 21*. Grao.

Savery, J. R. (2006). Overview of problem-based learning: Definitions and distinctions. *Interdisciplinary Journal of Problem-Based Learning*, *1*. https://doi.org/10.7771/1541-5015.1002. https://docs.lib.purdue.edu/ijpbl/vol1/iss1/3.

Srba, I., & Bielikova, M. (2015). Dynamic group formation as an approach to collaborative learning support. *IEEE Transactions on Learning Technologies*, *8*, 173–186. https://doi.org/10.1109/TLT.2014.2373374. https://ieeexplore.ieee.org/document/6963449.

Strom, P. S., & Strom, R. D. (2011). Teamwork skills assessment for cooperative learning. *Educational Research and Evaluation*, *17*, 233–251. https://doi.org/10.1080/13803611.2011.620345.

Van Den Beemt, A., Thurlings, M., & Willems, M. (2020). Towards an understanding of social media use in the classroom: A literature review. *Technology, Pedagogy and Education*, *29*, 35–55. https://doi.org/10.1080/1475939X.2019.1695657.

Vygotskij, L. S. (1979). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wang, R. (2022). Team diversity and team success in collaborative crowdsourcing. *Communication Studies*, *73*, 68–84. https://doi.org/10.1080/10510974.2021.2011355.

White, S. C. (2015). *Relational wellbeing: A theoretical and operational approach*. Bath papers in international development and wellbeing.

White, S. C. (2016). Introduction: The many faces of wellbeing (pp. 1–44). London: Palgrave Macmillan UK.

White, S. C. (2017). Relational wellbeing: Re-centring the politics of happiness, policy and the self. *Policy and Politics*, *45*, 121–136. https://doi.org/10.1332/030557317X14866576265970. https://bristoluniversitypressdigital.com/view/journals/pp/45/2/article-p121.xml.

Willis, S., Byrd, G., & Johnson, B. (2017). Challenge-based learning. *Computer*, *50*, 13–16. https://doi.org/10.1109/MC.2017.216.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, *330*, 686–688. https://doi.org/10.1126/science.1193147. https://www.sciencemag.org/lookup/doi/10.1126/science.1193147.

Wylde, D. (2013). *Post-Jungian personality theory for individuals and teams*. SYDROSE LP. https://books.google.es/books?id=-hAjngEACAAJ.

Zheng, Z., & Pinkwart, N. (2014). A discrete particle swarm optimization approach to compose heterogeneous learning groups. In *2014 IEEE 14th international conference on advanced learning technologies* (pp. 49–51). https://ieeexplore.ieee.org/document/6901395.