UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA
Dipartimento di Economia Marco Biagi

DEMB Working Paper Series

N. 229

Enhancing Student's Sense of belonging, performance and Inclusion: Exploring the Benefits of Team-Based Learning in University Classrooms

Chiara Tasselli[1], Tindara Addabbo[2]

October 2023

[1] University of Modena and Reggio Emilia, Department of Economics Marco Biagi
Email: chiara.tasselli@unimore.it
[2] University of Modena and Reggio Emilia, Department of Economics Marco Biagi
Email: tindara.addabbo@unimore.it

ISSN: 2281-440X online

Enhancing Student's Sense of belonging, performance and Inclusion: Exploring the Benefits of Team-Based Learning in University Classrooms

Chiara Tasselli, Tindara Addabbo

Department of Economics Marco Biagi,

University of Modena & Reggio Emilia

Abstract

Inclusive education can be defined as an educational approach proposing universities where all students can participate and feel part of the learning community. It is an educational philosophy and practice that aims to improve the learning and active participation of all students in a common educational context.

This paper analyses the benefits of introducing team-based learning (TBL) teaching methodology, an active learning method that promotes group work, in university classrooms.

The study compares two classes of macroeconomics with the same curriculum and teaching hours, but one class uses TBL while the other follows traditional teaching methods. The study goes beyond comparing students' performance at the individual level and assesses the potential of TBL to foster inclusion through the sense of belonging. The sense of belonging refers to an individual's subjective experience or perception of being connected, accepted, and included in a particular group, community, or social setting. It encompasses the feeling of being valued, respected, and integrated into a social context that in our case is the macroeconomic class/community.

Comparisons between groups were made using econometric and statistical analysis on a database composed of administrative sources and data collection originated through teaching activities which allow controlling of student socio-demographic characteristics and their academic careers. The econometric analysis started with a multivariate regression to estimate the TBL's effect on students' sense of belonging. Subsequently, it employed the propensity-score matching technique to match similar students and estimates: the average treatment effect (ATE), the average treatment effect on treated students (ATET) and the potential outcome mean (POM). Finally, it refined the estimates through the augmented inverse probability weighting. All techniques used confirm a positive and highly significant effect of treatment on students' sense of belonging.

In addition, multivariate analysis showed that those effects are higher for categories most at risk because they are more distant from the macroeconomic environment (females, low performers, repeat students, etc...).

The findings contribute to understanding the benefits of active learning approaches in terms of individual student success and promoting inclusivity for weaker groups in higher education.

Keywords: Innovative and inclusive education; cooperative learning; diversity; higher education; team-based learning; hurdle model.

JEL Classification: I23, I24, I25, J16

1. INTRODUCTION

The United Nations 2030 Agenda Sustainable Development Goal 4 "Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all" by focusing on equity and inclusion encourages the use of inclusive practices in all levels of education (UN, 2015). The Europe 2030 strategy identifies some crucial actions that European Union (EU) governments should pursue to boost growth and employment, especially for young people. Education and lifelong learning are listed among these essential goals. In this regard, the EU agenda encourages more inclusive and equitable access to education, a greater focus on addressing the learning crisis and teacher shortages, and increased support for education and training in emergencies and protracted crises (European Commission 2020).

The focus of this paper is on Italy. The system of higher education in Italy is undergoing a structural change guided by national reforms started in 2010 that address the quality of teaching and research, the efficiency of the system and seek to enhance Higher Education Institutions' (HEI) links with the ecosystems that are characterised in Italy by a high regional heterogeneity (OECD/European Union, 2019). Notwithstanding the effort to address the criticalities in the system of HEIs, Italy is still characterised by a very low percentage of people holding a tertiary degree. In 2022 the share of 25-34year-olds holding a tertiary degree in Italy was 29.2%, a percentage that is still much lower than the average for OECD and other EU countries (47.4%) (OECD, 2023, AlmaLaurea, 2023).

Considering the heterogeneity of Italian universities in terms of dimension, regional socioeconomic environment and higher education policies and with the intention of having an effect on the teaching strategies of the higher education institution analysed, this study will refer to a particular, medium-large university located in Northern Italy. This essay follows a strand of research that uses the survey method for the analysis of tertiary education in localised eco-environments (as for instance Kane *et al.* 2014). Since 2017, within the university analysed, a wider effort has been directed to the introduction of teaching strategies able to develop soft skills selected by external stakeholders (private and public organisations, enterprises, trade unions and their associations) belonging to the area where the HEI is located. As a main target of the project, the qualitative survey detected the development of problem solving and teamwork skills. The university then chose to fund the implementation of Team-based learning (TBL) considering its success in developing the two soft skills selected by the qualitative survey.

The methodology strategy at the core of the teaching practices analysed was developed by Michaelsen in the late 1970s and includes flipped-classrooms, problem-based learning, and activation methods to

develop active student engagement and specific soft skills. Though TBL is undergoing an increase in its implementation in Italy, this teaching methodology is still uncommon and the effort of evaluating it is still very limited in the Italian context.

The HEI project funded TBL training and support in its implementation with 16 lecturers in undergraduate and master's courses and 16 classes were involved as control groups. Together with a positive impact on the development of the relevant soft skills, the research on the outcomes of the implementation showed a positive contribution towards other students' learning outcomes such as student's performance and satisfaction with the teaching activities.

This essay aims to investigate another dimension of the impact of TBL methodology: its impact on students' inclusion. This is a dimension that is gaining increasing interest in the debate on the impact of universities in students' higher education experience and which is also at the centre of the chosen HEI policy objectives.

Studies on inclusion in higher education have analysed it in different dimensions/phases of the academic life from access to teaching and learning activities, assessment and extracurricular activities (Koutsouris, Stentiford, & Norwich, 2022). In this essay, attention will be paid to the impact of TBL on the teaching and learning environment, with special regard to a subject (economics) showing still very limited implementation of TBL (Allgood and McGoldrick, 2021; Cagliesi and Ghanei, 2022). The study will focus on students' sense of belonging, their learning experience and internal group dynamics, and the impact of TBL activity on specific minorities.

The essay then aims to enrich the body of knowledge on the impact of an active learning strategy (TBL) in two ways: by investigating its impact on a specific dimension (inclusion) and by focussing on a country (Italy) and on a subject (economics) where TBL is still underdeveloped.

Within inclusion, the essay will specifically address the sense of belonging that refers to students' experience of feeling included and accepted in their institution (Goodenow, 1993; Thomas, 2012).

The paper is structured as follows: section 2 opens with the literature review and section 3 describes the methods (experiment, variables and the empirical model). Section 4 follows with the results and section 5 expresses the conclusions and offers policy suggestions.

2. LITERATURE REVIEW

According to Espey (2022), active learning methods engage students in deeper thinking, drawing connections and applying concepts during class hours to create greater interest, learning and retention of the material. This is also confirmed by the Moriña and Orozco (2022) survey on Spanish university faculty members that shows that methodological strategies that are active and participatory for all students (like simulations, problem-based learning, flipped-classrooms…) are considered most effective in fostering inclusion. To provide an inclusive environment one often needs to combine more than one methodological strategy as stated by Tremblay-Wragg et al. (2022) and referred to by Lorenzo-Lledó et al. (2023).

The methodological strategy at the core of the teaching practices analysed (Team-based learning) as treatment in this paper, includes flipped-classrooms, problem-based learning, and activation methods and is guided by established pillars that have been shown to beneficially affect students' learning and skill enhancement and include retrieval practices, feedback and group working (Moore *et al.* 2020, Ruder *et al.* 2021, Simkins *et al.* 2021). The benefits of retrieval practices and feedback on the students' learning experiences have been highlighted by various studies (including Schell and Butler, 2018; Schwartz, Tsang, and Blair 2016; Simkins *et al.*, 2021).

TBL activities are conducted in small groups (five to six persons) that work together during the whole semester. Students' work in small groups encourages participants not to learn only from their own individual study or from the teacher but by interacting with peers in stable teams (Michaelsen, Watson, & Sharp, 1991; Opdecam *et al.* 2014). Groups are structured with the aim of maximizing homogeneity between groups and heterogeneity within each group. The formation of the groups plays a key role in both the treatment effectiveness and in terms of inclusiveness.

"TBL class structure requires repeated effortful retrieval practice by students throughout the semester and systematically provides timely, specific, and informative feedback to students on their learning during team-based activities. A TBL course is typically divided into 5 to 7 topical modules in a 15-week semester. Each module begins with a Readiness Assurance Process, consisting of out-of-class preparation by students (typically, reading a textbook chapter, watching videos, or listening to podcasts), an individual multiple-choice quiz (iRAT), and an in-class identical (or very similar) team-

based quiz (tRAT). The majority of time in each module is made up of a series of Application Exercises (AEs) that challenge students to apply and synthesize economic concepts and principles through the analysis of an increasingly complex set of real-world examples."

(Simkins *et al.*, 2021, p.233)

The essay deals with a special dimension of inclusion: sense of belonging regarding students' feeling of being included and accepted in their institution (Goodenow, 1993; Thomas, 2012). This can be considered as an important dimension of inclusion in HEIs and has been found to be also positively related to other dimensions such as students' engagement, self-confidence, academic achievement and retention (see Ahn and Davis, 2020 for survey of the literature).

3. RESEARCH METHODS

In this section we will illustrate the methodology followed to ascertain the impact of the chosen teaching strategy on students' sense of belonging.

The participants were students attending macroeconomics lectures in the academic year 2022/2023 as part of the bachelor course in business (n= 87) and marketing (n= 118).

They belong to the same cohort and attend Macroeconomics classes in their second year of a bachelor's degree but actually come from two different degree programs with different teachers.

We will address the whole sample to evaluate the impact of TBL on the sense of belonging. Meanwhile, we will focus on the treated subsample to investigate group dynamics and variations in their performance during the semester. This latter part is possible by taking advantage of all the information automatically produced during the activity (individual and group assessment, etc..).

Identification strategy and selection problems:

The sample initially consisted of all students involved, with a total of 253 students in the business degree program and 217 observations in the marketing group. Subsequently, a subset for observations was considered, focusing on those who had completed a questionnaire containing biographical and attitudinal information. This filtering process resulted in 131 observations for the business group and 120 observations for the marketing one. Moreover, the sample was further narrowed down for observation of those who also completed a postquestionnaire in which the dependent variable (sense of belonging) was detected.

Out of these, 87 observations from the first group satisfied the criteria for being included in the treated group: we only considered students who participated in a minimum of 5 out of the 6 TBL classes as part of the treated group.

To summarise, the final study sample consisted of 87 observations from the business degree program and 118 observations from the marketing one that met all the inclusion criteria.

Having to restrict the sample to only those who completed the questionnaire and those who attended TBL could create a sample bias towards attending students: lower attendance was of students who work, were enrolled in previous years and/or are less motivated, etc…

Regarding the sample restriction due to the questionnaire, we can exclude that it causes estimation problems since it is present – with the same dynamics – for both the treated and the control group.

To address the sample bias caused by eligibility treatment criteria, we gathered data on all pertinent confounding variables and incorporated them into our analysis when comparing the average exam scores between the treatment and control groups. Further details about how we addressed the issue

of endogeneity related to certain individual characteristics that could be correlated with treatment effects are reported in the section 3.3 dedicated to the empirical model.

Propensity score matching estimates both the average treatment effect (ATE) and the average treatment effect on the treated (ATET) by matching subjects based on their propensity scores. This technique allows us to impute the missing potential outcomes for each subject and adjust for covariates to estimate treatment effects.

Course design

For students in both the Team-Based Learning and the regular course, this was their first course on Macroeconomics. The courses covered the same topics using the same book and had a tutor who provided additional classes for practical exercises. The courses also had an identical number of credits and scheduled lessons: 36 lessons lasting one and a half hours but for the treated class, 6 of these lessons were used for the TBL activity.

All elements of the TBL implementation (including administrative and implementation ones) were extensively pre-tested since this practice had been carried out by the lecturer since the 2017/2018 academic year.

Twenty-seven groups were created using G(roup)Rumbler, an algorithm developed by Prof. Malcolm K. Sparrow in 2011, to ensure the heterogeneity of components within the group (Sparrow, 2011). The formation of the groups took into account student-related variables (gender, age, origin, type of secondary school attended, grades in Maths and Microeconomics, students' attitudes towards teamwork, and personal characteristics) in order to create groups or teams in a way that maximizes the diversity or heterogeneity of their members.

All teams consisted of 6 members (or 5 in one case) and were within the range of recommended team size (Michaelsen *et al.* 2004) and remained the same throughout the semester.

TBL implementation was guided by the 4-S framework (Carson *et al.* 2021; Clerici Arias 2021) that was core for this methodology: the exercise focused on a *significant problem*, the teams faced the *same exercises*, teams were required to make a *single choice* from 4 possible answers, and teams reported their choice *simultaneously*.

The teaching process also respected all the steps mentioned anticipated in the section on literature as components of the TBL structure. Students attended lectures having already studied the topics of the TBL session, and responded to the Readiness Assurance Tests containing multiple-choice questions: first individually (iRAT[1]) and then they answered the same questions in teams (tRAT[2]). Then the

[1] Where "*i*" state for individual.

[2] Where "*t*" state for team.

groups prepared to address the team application (tAPP). In the team application, students have to solve a topical problem using the concepts acquired during the study of the subject. The latter part implies a deeper level of reasoning as its solution is not provided in the study material: it must be found by deduction. In other words, the knowledge gained becomes a tool, not the solution. At the end of the lesson, the lecturers show all the exercises in the class (immediate feedback), but before doing this a debate is stimulated among groups on the tAPP solutions.

Finally, the groups are called to conduct peer evaluation in which each participant evaluates himself/herself and their teammates with regards to the knowledge of the study materials and the contribution and behaviour within the group. This part is usually carried out at the very end of the lessons or on the same day (with a strict deadline) in the event of running out of class time.

3.1 Description of Database and Variables

The variables collected concerned students' characteristics, their performance and attendance during the course, and their beliefs and satisfactions after the TBL activity.

A more detailed description follows:

DEPENDENT VARIABLES

The main dependent variable used in the comparison between the two classes was the "sense of belonging" (*S_belonging*). This variable is a continuous and composite indicator inspired by Good et *al.* (2012) that assessed participants' feelings of acceptance, connectedness, and membership within the course and the macroeconomics community. The variable was generated by combining the following students' feelings while they were attending the macroeconomics class:

1. I consider myself a member of the macro class / like I am part of the macro class.

2. I feel like an outsider. (R)

3. I feel accepted.

4. I feel respected.

5. I feel neglected. (R)

6. I feel valued.

7. I feel appreciated.

COVARIATES

MAIN COVARIATE OF INTEREST

Treat = corresponds to the treated group and – due to the adopted identification strategy – is also collinear with the bachelor's degree program to which the students belong.

OTHERS

Covariates are exclusively related to students' characteristics, behaviour and performance. No course-fixed effects are entered for two main reasons: firstly, the course designs are specular (cover the same topics, use the same books and have a tutor who provides additional classes), and secondly since only two classes are included, course-fixed effects are likely to be collinear or overlap with treatment. To make our estimates, we therefore preferred to exploit individual-level variation, leaving the treatment exclusivity of the class effect.

Students' background: demographic information such as gender (*Female*), *Domicile* and *Residence*; whether the student works (*Worker*) or whether he/she has had remunerative experience in the past (*Working_exp*). Moreover, in the descriptive statistics, two indicators of minority are also presented: *MinorityC* is a continuous indicator that adds up whether the student is female and/or has failed/not taken the microeconomics exam and/or is working and/or is overdue with taking the exam. Meanwhile, *MinorityD* is a dichotomic indicator indicating at least one of these conditions.

Students' behaviour/personality: if they are used to working in a team (*Team*), if they feel themselves to be *Leaders* rather than followers, if they are *Extroverts* rather than introverts and their in-group behaviour (*Advocancy_Role*; *Listening_Role*; *Mediation_Role*). If they have had experience of work (*Working_exp*) in the past and if Maths (*MathPass*) or Economics (*EcoPass*) are the subjects they are most passionate about.

Students' prior ability: prior ability was measured through the mark in microeconomics which is a subject taken in the first year relating to the skills required (four classes: *PendingExam* for those who have not yet taken or passed it; thereafter progressively *18to22*; *23to26*;*27to30*). Other indicators of prior ability are the secondary school attended (classical lyceum = *L_class*, scientific lyceum = *L_scie*, linguistic lyceum= *L_ling*, other lyceum = *L_other*, *technical* or *professional* secondary school) and whether students state that economics (*EcoSkill* dummy variable) or mathematics (*MathSkill* dummy variable) are the subjects on which their past education is predominantly based.

Students' observed ability: two are used as proxies of actual performance in the course: a dummy variable for passing the exam (*pass*) and a continuous variable concerning the grade (*mark*). In the event that students have taken the exam more than once, the grade considered is that of the last

attempt[3]. The descriptive statistics also contain the number of exam attempts (*attempts*) for students who have taken the exam.

As already mentioned, *Mark* is a continuous variable ranging from 17 (fail) to 30 (excellence). The grading scale is linear from 18 to 29, with 17 and 30 representing the lower and upper limits, respectively. This means that grades condense into the limits regardless of the level of fail or excellence. Since the descriptive statistics show that the lecturers in the two courses have different scales of measurement in assigning grades, in the analyses on the overall sample the grades were entered as standard deviation within the treated $((Z|_{T=1}) = (\frac{x_{i1}-\mu_1}{\sigma_1}))$ and the control $((Z|_{T=0}) = (\frac{x_{i1}-\mu_1}{\sigma_1}))$ groups. [1]

3.2 Descriptive statistics

3.2a -Whole sample

Descriptive statistics for the whole sample are displayed in Table 1. Treated students are 42% of the sample. As regards the performance in macroeconomics, as mentioned, we have a subsample of students who have taken the exam (157) and 85% of them passed it and the average pass mark was 25.12. Almost half of the sample is female (53%) and 6% of students were enrolled in previous academic years. 25% of students declared that they work. 78% of them are resident in the Emilia Romagna region. 75% of them have at least one characteristic (gender, low performer, etc.) that identifies them as belong to a minority (MinorityD).

As regards their past abilities, the majority came from a scientific lyceum (37%) or technical school (35%). 35% of them scored above 27 on the microeconomics exam and 17% have not yet taken or passed it[4]. Regarding their personal characteristics and attitudes: 66% of them have already had team activity experiences, 49% state that they are extroverts[5] and 54% that they are leaders (rather than followers). They declare that the subjects in which they have acquired more skills are economics (41%) and mathematics (27%) but only 9% of them state that mathematics is their favourite field against 53% for economics.

[3] Potentially equivalent to the verbalized grade unless the student failed or refused the grade on the last test.

[4] Because we are considering only students who took the macroeconomics exam in sessions close to the class period, we do not rule out having an upward bias in their abilities (approximated in the microeconomics grade). The latter due to the fact that we are excluding those who have not yet shown up at the macroeconomics exam sessions.

[5] The complementary antagonist option was shy.

Table 1 - Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Treat	205	.42	.5	0	1
Pass	157	.85	.35	0	1
Mark≥18	134	25.12	3.57	18	32
Mark	157	23.86	4.26	17	30
S_belonging	205	3.95	.7	1.71	5
Attempts	157	1.13	.36	1	3
MinorityD	205	.75	.44	0	1
MinorityC	205	1.04	.83	0	4
Female	205	.53	.5	0	1
Fuoricorso	205	.06	.24	0	1
Worker	205	.25	.43	0	1
Working_exp	205	.82	.38	0	1
High school					
L_Clas	205	.06	.24	0	1
L_Scie	205	.37	.48	0	1
L_Ling	205	.18	.39	0	1
L_Other	205	.03	.17	0	1
Technical	205	.35	.48	0	1
Professional	205	.01	.1	0	1
Mark in microeconomics					
PendingExam	205	.17	.38	0	1
18to22	205	.18	.39	0	1
23to26	205	.3	.46	0	1
27to30	205	.35	.48	0	1
Sud	205	.12	.32	0	1
Fuorisede	205	.15	.35	0	1
Team	205	.66	.47	0	1
Extrovert	205	.49	.5	0	1
Leader	205	.54	.5	0	1
SkillMajor	.	.	.	.	.
Mathematics	205	.27	.45	0	1
Italian	205	.1	.3	0	1
Law	205	.1	.3	0	1
Economics	205	.41	.49	0	1
Science	205	.04	.19	0	1
Other	205	.07	.25	0	1
PassMajor	.	.	.	.	.
Mathematics	205	.09	.29	0	1
Italian	205	.08	.27	0	1
Law	205	.23	.42	0	1
Economics	205	.53	.5	0	1
Science	205	.02	.14	0	1
Other	205	.05	.23	0	1
Domicile	.	.	.	.	.
Modena	205	.32	.47	0	1
Province of Modena	205	.24	.43	0	1
Reggio E. & Province	205	.24	.43	0	1
Other	205	.2	.4	0	1

Residence	.	.	.	.	.
Abruzzo	205	.01	.1	0	1
Calabria	205	.01	.12	0	1
Campania	205	0	.07	0	1
Emilia Romagna	205	.78	.42	0	1
Lombardia	205	.03	.18	0	1
Marche	205	0	.07	0	1
Molise	205	0	0	0	0
Piemonte	205	.01	.12	0	1
Puglia	205	.03	.18	0	1
Sicilia	205	.05	.23	0	1
Toscana	205	0	.07	0	1
Veneto	205	.01	.1	0	1
NA_other	205	.04	.19	0	1

If we analyse the sample into subgroups according to whether they belong to the treated or control group (Table 2) we can observe some differences. The control group has a significantly higher percentage (30 per cent) of individuals who belong to minorities. Almost the entire effect is due to a higher percentage of women in the control group.

A higher share of females is expected as the control group belongs to a less quantitative bachelor's degree. In fact, we also find consistency in that the control group has a significantly higher percentage of students from the linguistic lyceum and vice versa for the scientific one. In addition, the treated group had a better performance in microeconomics (as a quantitative subject) and a greater share of them stated that maths is the subject in which they are more skilled.

Lastly, if compared to the control group a very high percentage of students in the control group (63%) identify themselves as leaders rather than followers.

Regarding the performance in the macroeconomics exam: the control and treatment groups showed no statistical difference in either the probability of passing the exam (pass) or the grade itself if we take only verbalised scores (*Mark≥18*). Some statistical divergences arise when we consider them jointly (*Mark*[6]). Since the groups had different teachers, it could be that the yardstick adopted by them in giving grades was slightly divergent. As a result, in the data analysis section, for analyses involving the entire sample, we focused on observing standardised deviations within the class group, whereas we examined the grade's relative significance within the subgroups defined by the class. Another outcome variable that shows significant differences between the groups is the sense of belonging (*S_belonging*). This variable, unlike the grade which might be endogenous and have a spurious correlation with treatment (through the professor evaluation metrics), is intrinsically related to students' perceptions in the classroom. Hence, it is a good objective outcome for our study especially considering the fact that our focus is on inclusion.

[6] See variable description for more details.

Table 2 – Group comparison of personal characteristics and outcome variables

	(1) CONTROL		(2) TREAT		(3) T-TEST ($\overline{x_C} - \overline{x_T}$)	
	mean	**sd**	**mean**	**sd**	**b**	**t**
S_Belonging	3,61	0,60	4,42	0,53	-0,81***	(-10,18)
Pass	0,90	0,30	0,82	0,39	0,08	(1,52)
Mark≥18	25,62	3,07	24,68	3,92	0,94	(1,56)
Mark	24,74	3,88	23,15	4,44	1,59*	(2,40)
Attempts	1,09	0,28	1,17	0,41	-0,09	(-1,57)
MinorityD	0,87	0,33	0,57	0,50	0,30***	(4,84)
MinorityC	1,24	0,77	0,69	0,67	0,55***	(5,43)
Female	0,64	0,48	0,37	0,49	0,28***	(4,05)
Fuoricorso	0,06	0,24	0,06	0,23	0,00	(0,06)
Worker	0,31	0,46	0,17	0,38	0,13*	(2,25)
Sud	0,14	0,35	0,08	0,27	0,06	(1,45)
Fuorisede	0,19	0,39	0,09	0,29	0,09*	(1,98)
LicClas	0,07	0,25	0,06	0,23	0,01	(0,30)
LicSC	0,27	0,45	0,51	0,50	-0,23***	(-3,46)
LicLING	0,27	0,45	0,06	0,23	0,21***	(4,44)
AltroLIC	0,03	0,18	0,02	0,15	0,01	(0,47)
Tecn	0,35	0,48	0,34	0,48	0,00	(0,04)
Prof	0,01	0,09	0,01	0,11	-0,00	(-0,21)
Working experience	0,81	0,39	0,84	0,37	-0,03	(-0,48)
PendingExam	0,23	0,42	0,09	0,29	0,14**	(2,75)
18to22	0,24	0,43	0,10	0,31	0,13**	(2,61)
23to26	0,35	0,48	0,24	0,43	0,11	(1,66)
27to30	0,19	0,39	0,56	0,50	-0,38***	(-5,84)
Team	0,60	0,49	0,75	0,44	-0,15*	(-2,23)
Extrovert	0,48	0,50	0,49	0,50	-0,01	(-0,16)
Leader	0,63	0,49	0,43	0,50	0,20**	(2,90)
Advocancy_Role	0,25	0,43	0,18	0,39	0,06	(1,07)
Listening_Role	0,19	0,39	0,26	0,44	-0,08	(-1,31)
Mediation_Role	0,55	0,50	0,55	0,50	-0,00	(-0,01)
EcoSkill	0,53	0,50	0,25	0,44	0,28***	(4,27)
MathSkill	0,18	0,38	0,40	0,49	-0,22***	(-3,53)
EcoPass	0,57	0,50	0,47	0,50	0,10	(1,37)
MathPass	0,10	0,30	0,08	0,27	0,02	(0,52)
Emilia Romagna region	0,69	0,46	0,89	0,32	-0,19***	(-3,47)
N	**118**		**87**		**205**	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Finally, with reference to the covariates, divergences between groups are not a problem, but they must be taken into account when estimating the treatment effect. In fact, they imply that we cannot estimate the average treatment effect (ATE) simply by taking the difference between the sample means for treated and control students, because there are covariates related to potential outcomes and treatment StataCorp.(2021). Moreover, we add post-estimation tests and diagnostic statistics to verify the balance of covariates as a specification check of our model. All these concepts are expanded in the following section dedicated to the methodology and the empirical model.

3.3b -Treated subsample

In light of the evidence highlighted by the descriptive analyses, assessing performance across the academic courses is not suggested as this may suffer from endogeneity. However, we could make use of the extensive administrative data generated during TBL activities as a robust and consistent indicator for a useful insight on students' performances.

These analyses exclude the counterfactual but include a larger sample of TBL participants. Because the analyses focus exclusively on within-group performance trends, the requirements of meeting the minimum treatment dosage and filling out all the questionnaires are waived. This allows the treated sample size to be reduced from 142 to 120 students depending on the frequency of TBL sessions.

Table A1 in the appendix – for all 27 teams – presents individual and group results for each of the six TBL sessions.

For individual performance (I-Rat), it includes for each group the minimum (worst performers), maximum (best performers) and group mean values, along with indicators of intergroup variance (variance and standard deviation). Conversely, for T-Rat not all these indicators are present as the group joins together and generates a single indicator.

Two key initial evidences emerge from this table:

(i) The results of the team when it is together are always higher than the average of the group where each member works individually $TRat_g > \overline{IRat_{ig}}$ with i (1,.. 6) ; g (1,.. 27)

(ii) The group working together usually also achieves higher scores than the best performer when working alone $TRat_g > \max_{Irat_i \in G_g}(IRat_i)$ with i (1,.. 6) ; g (1,.. 27)

Table A2 in the appendix, offers a further insight on these results.

Cells report the gains of working in a group. These are measured through the distance of the Trat score from:

a) The Group Average Individual Performance ($TRat_g - \overline{IRat_{ig}}$) with i (1,..6) ; g (1,..27);
b) The Best performer ($TRat_g - \max_{Irat_i \in G_g}(IRat_i)$ with i (1,..6) ; g (1,..27);
c) The Worst performer. ($TRat_g - \min_{Irat_i \in G_g}(IRat_i)$ with i (1,.. 6) ; g (1,.. 27).

The visualization is accompanied by colour-coding to enhance the clarity of gap visualization. The darker the green, the higher the performance of the collaborative group compared to the reference unit. The *worst performers* tend to benefit the most from group activities (Trat). At the same time, the score of individuals working in a team (Trat) consistently exceeds the group average when individual members work independently ($\overline{IRat_{ig}}$). Finally, in addition, the gains of *top performers* are consistently positive for most groups. Occasionally, in some groups, a *best performer* had an I-

Rat score higher than T-Rat, but we are talking about one (or two) out of six TBL sessions. In the remaining cases, it is the group working together that perform better. Some authors call the percentage by which the group performance score was higher (or lower) than the group's best member as the Group Added Value (Watson, Michaelsen, Sharp 1991).

Finally, table 3 shows the trend analysis in the average performance scores for individuals (Column 1) and teams (Column 2) over time. The third column (Column 3) illustrates the mean standard deviation across the 27 groups, providing insights into the consistency or variability within these groups. It can be seen that over time both individual and group average scores have an increasing trend. This is a sign that, over the course of the activity, students improve in solving the exercises both alone and in groups. The only observation out of trend is the one related to the last TBL, but it concerned a part of the course that students found particularly difficult[7]. In addition, there is a decreasing trend in heterogeneity of performance within the course, although not entirely linear. This trend indicates a positive aspect of inclusiveness, as it demonstrates how TBL practical experience tends to mitigate differences among participants' performances.

Table 3 – TBL Performance - trend analysis

Session	Obs	(1) Mean I-Rat	(2) Mean T-Rat	(3) Mean SD(Irat) intra group
TBL1	142	5.33	8.88	1.88
TBL2	140	6.28	9.1	1.66
TBL3	141	7.34	9.91	1.99
TBL4	133	7.89	9.78	1.44
TBL5	120	8.81	9.98	1.39
TBL6	128	7.48	9.66	1.54

3.3 Empirical Model

The data analysis section (as well as the empirical model) consists of two subsections: one that uses both courses while the second focuses only on the business course.

In the first, we used the counterfactual technique to estimate the effect of TBL on the sense of belonging while in the second we analysed the data generated during the teaching activity to observe the change in individual and group performance over time.

Paragraph 4 opens with an explanatory analysis in which the relationships between students' characteristics and some outcome(s) of interest are observed through person correlation and the point biserial correlation. Then, the regression in our analyses is structured as reported in equation 1, where

[7] In the last questionnaire there was an item regarding the most difficult part in the course.

subscripts i and c denote respectively individual and courses, Y_{ic} denotes the *Sense of belonging;* T_{ic} represents a dummy variable indicating if the students participated to the TBL activity; X_{ic} represents a vector of characteristics of the individual i within the course c, and x_{ic} represents a subvector of the characteristics contained in X respectively (gender and low past performance). $\beta_1;\ \beta_2;\ \beta_3;\ \beta_4$ represent the coefficients to be estimated: in particular β_1 reveals the treatment effect and β_3 (coefficient of the interaction term) explores the effect of treatment on most vulnerable groups. Finally, the term ε_{ic}, represents the error terms. For the subsample of the population that has already taken the macroeconomics exam, we have also entered the actual performance indicators (γ_{ic} in Equation 2) for checking whether ability in the subject played a role in the sense of belonging to the macroeconomics community.

(1) $Y_{ic} = \alpha + \beta_1 T_{ic} + \beta_2 x_{ic} + \beta_3 (T_{ic} \cdot x_{ic}) + \beta_4 X_{ic} + \varepsilon_{ic}$ with i (1,.. 205) ; c (1, 2)

(2) $Y_{ic} = \alpha + \beta_1 T_{ic} + \beta_2 X_{ic} + \beta_3 (T_{ic} \cdot x_{ic}) + \beta_4 X_{ic} + \beta_5 \gamma_{ic} + \varepsilon_{ic}$ with i (1,.. 157) ; c (1, 2)

Results of equation (1) and (2) are displayed in table 4 in the data analysis section and are repeated also by course subgroups[8].

Moreover, as mentioned in the previous section, the evidence arising from the descriptive statistics and from the data analysis suggests to us to make a more in-depth analysis and check for the endogeneity of some individuals' characteristics which could be related to treatment.

We have computed the treatment effect also by using i) the propensity-score matching and ii) the augmented inverse-probability weighting.

The *propensity score matching* estimates the average treatment effect (ATE) and average treatment effect on the treated (ATET) from observational data by propensity-score matching. The latter imputes the missing potential outcome for each subject by using an average of the outcomes of similar subjects that receive the other treatment level. The similarity between subjects is based on the probability of receiving treatment given $X = x$, known as propensity scores (equation 3):

(3) $PS(T_i \mid X_i) = e(X) = Prob\{T = 1 \mid X = \alpha + \beta_1 X_{ic} + \varepsilon_i\}$

The treatment effect is computed by taking the average of the difference between the observed and potential Sense of Belonging for each subject. Rosenbaum and Rubin (1983) demonstrated that adjusting for covariates x_i (and using the probability of treatment to perform the adjustment) is sufficient to estimate the effects. The first limit of this methodology is that in our case there are some

[8] $Y_{i|c=0} = \alpha + X_i\beta_1 + \varepsilon_{1i}$
$Y_{i|c=1} = \alpha + X_i\beta_1 + \varepsilon_{1i}$
$Y_{i|c=0} = \alpha + X_i\beta_1 + \gamma_i\beta_2 + \varepsilon_{1i}$
$Y_{i|c=1} = \alpha + X_i\beta_1 + \gamma_i\beta_2 + \varepsilon_{1i}$

covariates related both to potential outcomes and treatment (gender, micro scores Eco|math skills and eco passion).

The *augmented inverse-probability weighting* (AIPW) estimates the effect by using weighted regression models which in addition to taking into account the probability of treatment (as the propensity score in equation 3) it also exploits another model to predict outcomes.

AIPW estimators employ a three-step approach to estimate treatment effects:

In the first step, the inverse-probability weights are computed through the parameters of the treatment model (Eq. 3).

$$\textbf{(4)}\quad W_i = \frac{T_i}{PS_i} + \frac{1-T_i}{1-PS_i}$$

For treatment independent variables, we use the main predictors that Table 2 shows in the descriptive statistics as unbalanced among treated and control groups (gender and prior competencies such as the type of secondary school diploma, excellence in microeconomics and skills in mathematics).

In the second step, separate regression models for the outcome are estimated for each treatment level, and treatment-specific predicted outcomes for each student are obtained.

$$\textbf{(5)}\quad E(Y|T_i\,, X_i) = \alpha + X_i\beta_1 + T_i\beta_2 + \gamma_i\beta_3 + \varepsilon_i$$

Finally, in the third step, weighted averages of treatment-specific predicted outcomes are calculated, where the weights are the inverse probability weights calculated in the first step.

$$\textbf{(6)}\quad ATE = \frac{1}{N}\sum_{i=1}^{N} W_i \cdot E(Y|T_i\,, X_i)$$

For both equations 3 and 5, the full specification model remains consistent with the one of the regression model, but with the omission of the actual performance, the domicile and the residence fixed effects and the simplification of group attitudes. The omission of these variables is justified by the non-significance of their coefficients and their low contribution in increasing the adjusted R^2. Moreover, its omission allows a higher sample size to be reached.

Due to the fact that this model incorporates both the outcome and treatment probabilities, it is known as being “doubly robust” and more efficient (StataCorp 2021).

This model has several other positive properties in terms of goodness of estimate, also including tolerance for misspecification: in fact, it is sufficient that only one among the models of treatment status and outcome prediction is correctly specified to be consistent (asymptotically unbiased).

Results of the treatment effect via these two techniques based on the propensity score association are displayed in tables 6 and 7.

4. RESULTS

Table 4 – Correlation between the main outcome and covariates

	Sense of Belonging		
	(a) ALL	(b) CONTROL	(c) TREATED
Mark[1]	0,52***	0,21	-0,06
Trattato	0,58***	NA	NA
Attempts	-0,07	-0,18	-0,15
Female	-0,17*	-0,08	0,09
fuoricorsoESSE3	-0,14*	-0,30**	0,01
Worker	-0,07	0,06	-0,05
Sud	0,01	0,04	0,18
Fuorisede	-0,05	-0,03	0,17
LicClas	-0,09	-0,08	-0,13
LicSC	0,16*	0,03	0,02
LicLING	-0,11	0,03	0,15
AltroLIC	-0,04	-0,13	0,17
Tecn	-0,02	0,03	-0,10
Prof	0,01	-0,01	0,00
EsperienzeLavoro	0,01	0,11	-0,17
NonSostenuto	-0,21**	-0,15	-0,07
tra18e22	-0,08	0,05	-0,05
tra23e26	-0,07	0,06	-0,10
tra27e30	0,29***	0,03	0,15
Team	0,08	0,02	-0,05
Estroverso	0,01	0,03	-0,02
Leader	-0,12	0,06	-0,08
gruppo_mieIdee	-0,06	-0,10	0,10
gruppoAscolto	0,09	0,03	0,06
gruppoMediazione	-0,02	0,08	-0,15
EcoSkill	-0,08	0,21*	-0,08
MathSkill	0,14	-0,14	0,15
EcoPass	0,03	0,17	0,02
MathPass	-0,01	-0,06	0,15

Notes:
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
[1] When we address the entire sample, grades are standardised within courses.
For continuous Items Pearson Correlation Coefficient is observed, whereas the association between dichotomic and continuous variables is computed through the Point Biserial Correlation.

Table 4 reports the measures of association between the *Sense of belonging* and students' covariates. Analyses are repeated for a) the group as a whole b) the control group and c) the treated one.

Measures of associations are computed through the Pearson Correlation Coefficient when two continuous variables are involved and through the Point Biserial Correlation for the association between dichotomic and continuous variables.

In the overall sample (column a), the first interesting point is that there is a strong and significant positive relationship[9] between grades and the sense of belonging to the macroeconomics community/class. Another interesting fact is that the treatment (TBL participation) and the Sense of belonging have the strongest association in the table. In column a we can see positive association between the sense of belonging and the performance in microeconomics. Other associations are the positive one regarding students from scientific secondary schools, and the negative one regarding female students and students from previous years.

In the control group (column b), the negative association also relates to students from previous years and the positive one to those who think that economics is their core educational background. Meanwhile for the treated group (column c), there emerges a significant association with covariates. Regardless of starting characteristics, all students in the business course seem to have a greater sense of belonging, potentially caused by TBL practice.

Table 5 shows the results of equations (1) and (2). Both models are also repeated for subsamples of the treated (column C) and control groups (column D), with the omission of the treatment variable and hence its interactions. The distinction between the two models is that in the second we also control for ability in Macroeconomics (entered as the interaction between the probability to pass and the observed mark in Macroeconomics).

In both models, treatment has a positive and significant effect on students' sense of belonging namely + 11% (0.552 on a 5-point scale) in the first model and + 21% (1.051) in the second.

We see that for the control group (models 1b), females suffer from a lower sense of belonging to the field which is usual in literature as females are seen (and self-viewed) as distant and unsuited to quantitative subjects. This is due to common, ingrained societal thoughts (such as stereotypes) that have a poor empirical basis. Also, other students that could be less confident in their ability in the subject (students from previous years and those who have not yet taken the microeconomics exam) demonstrate a lower sense of belonging. Conversely, those who have a strong economic background have a greater sense of belonging. All these evidences together potentially result in a harmful cycle of polarization. In this cycle, those who already have an affinity for the subject matter tend to become even more aligned with it, while individuals who are less in line with the subject matter (potentially

[9] further analysis or investigation of the literature should be done on the causal relationship as it is unclear whether being good at economics causes a greater sense of belonging or a high sense of belonging has a positive impact on performance (perhaps through effort and/or self confidence).

more vulnerable) may come to believe that it is not a good fit for them. The sign of the association is also confirmed by including the grade (model 2b), although slightly weakening in significance, but probably also due to the reduction in sample size.

On the other side, in the treated group (models 1c and 2c) there is non-significance of any coefficient regarding demographics or performance. In addition, the R^2 and the adjusted R^2 are very low, a sign that the used covariates are not good predictors to explain the variance in sense of belonging. It is important to remember that Sense of Belonging was measured once the treated group had received treatment, so it may have had an impact in smoothing out the effects of the covariates.
Reasoning in terms of inclusion, these are positive and relevant pieces of information. In fact, the greater (or lower) *Sense of Belonging* to the macroeconomy community does not seem to depend on personal characteristics (gender, skills, working status, etc..) or prior education on current observed ability. This places all class members on the same level without distinction or ghettoing and could help in overcoming and smoothing polarization.

Table 5 – Regression on sense of Belonging

		1			2	
	(1a)	(1b)	(1c)	(2a)	(2b)	(2c)
	All	Control	Treated	All	Control	Treated
Treat	0,552***			1,051**		
	(0,147)			(0,380)		
Treat·Female	0,311			0,187		
	(0,176)			(0,218)		
Treat·fuoricorso	0,751*			0,607		
	(0,354)			(0,498)		
Treat· PendingExam	0,097			-0,516		
	(0,266)			(0,348)		
Female	-0,219	-0,269*	0,095	-0,142	-0,226	0,041
	(0,124)	(0,130)	(0,152)	(0,172)	(0,183)	(0,152)
Fuoricorso	-0,659**	-0,731**	0,235	-0,586	-0,329	0,054
	(0,230)	(0,232)	(0,306)	(0,395)	(0,479)	(0,371)
PendingExam	-0,246	-0,349*	-0,023	0,368	0,161	-0,108
	(0,147)	(0,156)	(0,259)	(0,265)	(0,303)	(0,259)
Fuorisede	0,503	0,424	0,441	0,329	0,280	0,554
	(0,708)	(0,864)	(0,328)	(0,733)	(0,948)	(0,329)
L_Scie	0,080	0,140	-0,138	-0,097	-0,009	-0,148
	(0,119)	(0,147)	(0,263)	(0,151)	(0,239)	(0,261)
Techinical	0,011	0,063	-0,129	-0,086	0,028	-0,121
	(0,121)	(0,142)	(0,298)	(0,162)	(0,225)	(0,298)
Professional	-0,162	-0,281	-0,408	-0,140	-0,120	-0,508
	(0,424)	(0,585)	(0,685)	(0,461)	(0,669)	(0,675)
Working exp	0,008	0,183	-0,156	0,008	0,027	-0,119
	(0,119)	(0,153)	(0,218)	(0,144)	(0,212)	(0,216)
27to30	0,056	-0,050	0,144	0,047	-0,262	0,272

	(0,100)	(0,158)	(0,151)	(0,110)	(0,202)	(0,162)
Team	-0,025	-0,003	-0,064	-0,065	-0,064	-0,144
	(0,093)	(0,119)	(0,168)	(0,112)	(0,174)	(0,170)
Extrovert	-0,011	-0,110	0,100	0,045	0,013	0,073
	(0,095)	(0,122)	(0,176)	(0,114)	(0,181)	(0,174)
Leader	-0,007	0,114	-0,125	-0,061	-0,047	-0,151
	(0,095)	(0,132)	(0,165)	(0,117)	(0,216)	(0,164)
EcoSkill	0,195	0,318*	0,104	0,306*	0,484	0,180
	(0,117)	(0,153)	(0,228)	(0,149)	(0,257)	(0,229)
EcoPass	0,042	-0,106	0,139	0,015	-0,172	0,070
	(0,100)	(0,143)	(0,162)	(0,117)	(0,222)	(0,163)
MathSkill	-0,026	-0,259	0,165	0,107	-0,158	0,263
	(0,124)	(0,193)	(0,198)	(0,147)	(0,311)	(0,202)
MathPass	0,204	0,181	0,345	0,306	0,239	0,209
	(0,165)	(0,236)	(0,290)	(0,201)	(0,422)	(0,305)
Advocancy_Role	0,375	0,232	0,201	0,448	0,151	0,235
	(0,420)	(0,428)	(0,176)	(0,430)	(0,461)	(0,174)
Listening_Role	0,513	0,663	0,131	0,567	0,796	0,143
	(0,424)	(0,440)	(0,159)	(0,435)	(0,489)	(0,157)
Mediation_Role	0,394	0,421		0,390	0,339	
	(0,416)	(0,419)		(0,429)	(0,456)	
Pass				6,784	0,131	0,070
				(5,659)	(0,421)	(0,277)
Pass·Z(markhurdle)				-1,519		
				(1,279)		
markHurdle					0,041	-0,043
					(0,031)	(0,023)
Domicile FE	YES	YES	YES	YES	YES	YES
Residence FE	YES	YES	YES	YES	YES	YES
Constant	2,373*	2,117*	4,245***	2,236	1,257	5,086***
	(0,945)	(1,060)	(0,336)	(1,134)	(1,384)	(0,531)
N	205	118	87	157	70	87
Adjusted R^2	0,368	0,192	-0,155	0,271	0,187	-0,117
R^2	0,482	0,420	0,181	0,444	0,564	0,234

Notes:

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Nonscientific lyceum variables omitted for collinearity in secondary school background (base for *L_scie; technical; professional*)

Mark in column 2a is standardised at class level, whereas in column 2b and 2c it is the relative value. For that reason, the interaction between pass and Mark is omitted for the collinearity in column 2b and 2c, whereas the same for mark in column 2a. But the functional form is the same for all sample groups.

Table 6 shows the treatment effect computed through the propensity score (equation 3) matching (PSM). The average treatment effect (ATE) measures the effect on the whole population and has a positive and significant outcome (0.783). It could be interpreted as follows: the average "Sense of belonging" if all students participated in TBL would be 15.6% (0.783 on 5) higher if all students had participated in TBL. The average treatment effect on the treated (ATET), indeed focuses only on

individuals actually receiving treatment, thus providing a more specific estimate for this subgroup. We can see that the effect decreases in magnitude (0.696), but remains positive and highly significant. The interpretation of the coefficient is that the effect of TBL on those treated amounted to an increase of 14% in their sense of belonging. This result is consistent with what emerged from the regressions in Table 4; in fact, the counterfactual group is also characterised by a higher percentage of students subject to a lower sense of belonging who are also those who better react to treatment. Appendix A3 provides the Kernel density plot of the propensity score both for the ATE and the ATET model.

Table 6 – Treatment effect through the propensity score matching

Dep.Var: sense of belonging		
	Coefficients	[95% conf. interval]
ATE	0.783***	[0.59 0.973]
Treat (1 vs 0)	(0.977)	
ATET	0.696***	[0.380 1.013]
Treat (1 vs 0)	(0.161)	
N	***205***	

Notes: *t* statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Estimator: propensity-score matching; Outcome model: matching
Treatment model: logit; Robust Abadie–Imbens standard errors with # matches
Caliper (0.09[10])

Due to the nature of starting differences between counterfactual and treated differences, we introduced several specification model checks.

Panel A and B in Appendix A3 shows diagnostic statistics to check for covariate balance for this model.

The more the standardised differences are all close to zero and the variance ratios are all close to one, the more the covariates are considered balanced. We can see that for both models (ATE in panel A and ATET in panel B), the matched values (b) are better than the Raw one (a), but not optimal. In fact, some covariates have a variance ratio that is slightly offset from one.

Thus, we introduced some improvements in estimates in the following model (table 6) illustrated in the equations 3 to 6.

Table 7 shows the results for the AIPW which combines the estimates from the propensity score model and the outcome model to provide a doubly robust estimate of the treatment effect.

Panel C in Appendix A3 concerning the specification check confirms that this second model is optimal and preferred with respect to those in table 5; as for the matched values, it shows both

[10] Only one observation exceeded the caliper(0.08).

standardised differences close to zero and a variance ratio close to one. Moreover, we tested whether the model balances all eight covariates used and we cannot reject the null hypothesis that the covariates are balanced (H0: Covariates are balanced; Prob > chi2 = 0.63). Through this test, we received the confirmation, in a model-based approach, that the propensity scores were correctly specified (Imai & Ratkovic 2014).

In this model the ATE (column 1) is significant and has a magnitude of 0.896: This indicates that, if all students had participated in TBL, the average sense of belonging would be 0.896 more than the average of 3.615 (column 2) that would occur if none of the students had attended. This signifies an increase of nearly 25 percent over the baseline conditions.

Column 3 in table 7 represents the linear regression coefficients for the untreated potential-outcome equations, whereas column 4 contains the coefficients for the treated potential outcome equation.

Finally, the last column (5) in table 7 contains the coefficients that are used in the probit model to predict treatment status: the results are consistent with previous and descriptive analyses: females are less likely to be treated and those who had a high grade in microeconomics are more likely to be treated.

Table 7 – Treatment effect through the augmented Inverse Probability Weighting

	(1) ATE	(2) POmean	(3) OME0	(4) OME1	(5) TME1
ATE	0,896***				
	(0,101)				
Treat = 0		3,615***			
		(0,060)			
Female			-0,198	0,113	-0,665**
			(0,124)	(0,128)	(0,207)
fuoricorsoESSE3			-0,739*	0,185	
			(0,318)	(0,250)	
Worker			0,027	-0,073	
			(0,129)	(0,152)	
Fuorisede			-0,014	0,407*	
			(0,135)	(0,174)	
High school					
L_Class[Base]					-0,031
					(0,405)
L_Scie			0,174	0,334	0,375
			(0,199)	(0,259)	(0,231)
L_Ling			0,139	0,820***	
			(0,208)	(0,234)	
L_Other			-0,230	0,809**	
			(0,286)	(0,255)	
Tecnical			0,054	0,336	
			(0,247)	(0,257)	
Professional			-0,130	0,124	
			(0,303)	(0,359)	
Working exp.			0,171	-0,129	

	(0,139)	(0,146)	
Micro Score			
Pending exam[Base]			
18to22	0,383*	-0,061	0,439
	(0,162)	(0,248)	(0,373)
23to26	0,277	0,003	0,525
	(0,158)	(0,246)	(0,325)
27to30	0,305	0,189	1,393***
	(0,172)	(0,235)	(0,321)
Team	-0,013	-0,048	
	(0,113)	(0,146)	
Estroverso	-0,120	0,114	
	(0,112)	(0,140)	
Leader	0,054	-0,075	
	(0,120)	(0,128)	
EcoSkill	0,265	0,093	
	(0,137)	(0,180)	
EcoPass	0,001	0,145	
	(0,118)	(0,116)	
MathSkill	-0,122	0,163	0,382
	(0,137)	(0,181)	(0,230)
MathPass	0,090	0,346*	
	(0,202)	(0,169)	
Constant	3,200***	3,824***	-0,819**
	(0,240)	(0,316)	(0,303)
N	***205***		

Notes: *t* statistics in parentheses $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Estimator: augmented IPW

Outcome model: linear by ML

Treatment model: probit

POmean = Potential outcome mean

OME0 = Outcome model parameter estimates for control

OME1= Outcome model parameter estimates for treated

TME1= Parameter estimates for treatment model

5. CONCLUSIONS

When analysing the positive impacts of an educational methodology, the focus solely on grades is limiting. This research uses the Sense of Belonging as the main dependent variable. In addition to being closely related to the concept of inclusion and representing student sentiment, this has a positive and indirect impact on performance. When the sense of belonging is low, individuals may choose to drop out of the field, become less committed and/or have lower self-confidence – even if achievements remain high – in order to pursue studies and professional goals within a different discipline where the sense of belonging can take root more effectively. The positive relationship between the sense of belonging and actual performance is confirmed by our correlation analysis for the overall sample and remains positive for the control group though it loses statistical significance for the relationship of the treated group. This could be a sign of a new variation which affects all the treated group and has not yet had an impact on their performance or may be a sign of the loosening of the link between low membership and low performance.

Descriptive statistics show, as expected, that the treated and control groups are heterogeneous in terms of characteristics, which motivates the estimation methodologies employed. The positive impact of treatment already emerges from the regression models: treatment results in an 11 percent increase in the sense of belonging, rising to 21 percent when corrected for the grade obtained in the exam. Furthermore, in this model, we can see that in the control group, personal characteristics that determine minority status have a negative impact on the sense of belonging, while in the treated group no personal characteristics can explain it. The estimates conducted through the propensity score aggregation shows both positive and significant average treatment effect (ATE) and average treatment effect on treated (ATET). The former is greater and this is coherent with what emerged from the regressions in Table 4, where the control group had a higher proportion of students with a lower sense of belonging who responded better to the treatment.

Finally, the last estimates, concerning the most precise model: the average treatment effect is always higher + 0.986 compared with not participating in TBL activity (3.615).

These results call for a wider use of TBL as a teaching strategy, since it positively contributes to students' sense of belonging with a positive impact on inclusion and students' performance that is amplified by the observed minority students' achievements.

6. Appendix

Appendix A1 – Individual and group performances

	Stats	TBL1	TBL2	TBL3	TBL4	TBL5	TBL6
GROUP 1	**Trat**	**8,4**	**9,4**	**10**	**10**	**10**	**9,7**
	IratMean	5,6	7	8,3	8,5	9,2	7,7
	Irat Max	7,5	8,8	10	10	10	8,8
	Irat Min	3,8	6,3	3,8	7,5	7,5	6,3
	Irat Var.	3	1,3	6,7	1,5	1	0,9
	Irat SD	1,7	1,1	2,6	1,2	1	0,9
GROUP 2	**Trat**	**7,5**	**8,8**	**10**	**10**	**10**	**10**
	IratMean	4,6	6	8,3	8,3	10	8,8
	Irat Max	6,3	8,8	10	10	10	10
	Irat Min	2,5	1,3	5	2,5	10	7,5
	Irat Var.	2,9	7,1	3,6	10,6	0	1
	Irat SD	1,7	2,7	1,9	3,3	0	1
GROUP 3	**Trat**	**8,8**	**10**	**10**	**10**	**10**	**8,8**
	IratMean	6	6,8	7,5	8,1	9,2	7,5
	Irat Max	7,5	8,8	10	8,8	10	8,8
	Irat Min	3,8	5	2,5	6,3	7,5	6,3
	Irat Var.	1,9	3,6	8,6	1,6	2,1	2,1
	Irat SD	1,4	1,9	2,9	1,3	1,4	1,4
GROUP 4	**Trat**	**9,4**	**10**	**10**	**10**	**10**	**10**
	IratMean	5,6	6,3	8,4	8,8	10	9,4
	Irat Max	7,5	8,8	10	10	10	10
	Irat Min	2,5	2,5	6,3	7,5	10	8,8
	Irat Var.	4,7	11,7	2,5	1,6	0	0,8
	Irat SD	2,2	3,4	1,6	1,3	0	0,9
GROUP 5	**Trat**	**9,1**	**8,8**	**10**	**9,4**	**10**	**9,7**
	IratMean	5,3	7,5	6,5	6,8	10	7,5
	Irat Max	6,3	8,8	10	7,5	10	8,8
	Irat Min	5	6,3	3,8	3,8	10	5
	Irat Var.	0,4	0,8	7,3	2,8	0	3,1
	Irat SD	0,6	0,9	2,7	1,7	0	1,8
GROUP 6	**Trat**	**8,1**	**9,4**	**9,4**	**9,1**	**10**	**9,4**
	IratMean	4,8	4,8	5	7,1	8,3	6,7
	Irat Max	6,3	7,5	8,8	8,8	10	8,8
	Irat Min	3,8	1,3	0	5	2,5	3,8
	Irat Var.	1,5	5,3	8,8	1,7	10,6	3,5
	Irat SD	1,2	2,3	3	1,3	3,3	1,9
GROUP 7	**Trat**	**8,8**	**10**	**10**	**9,7**	**10**	**9,7**
	IratMean	4,2	5,8	7,9	8,1	9	6,3
	Irat Max	6,3	7,5	10	8,8	10	8,8
	Irat Min	1,3	3,8	5	6,3	7,5	3,8
	Irat Var.	2,9	3,5	4,2	1,1	1,5	3,9
	Irat SD	1,7	1,9	2	1	1,2	2

	Stats	TBL1	TBL2	TBL3	TBL4	TBL5	TBL6
GROUP 8	**Trat**	**8,8**	**10**	**10**	**10**	**10**	**9,4**
	IratMean	4,8	6,8	6,3	8,1	8,5	7,5
	Irat Max	7,5	8,8	8,8	10	10	10
	Irat Min	2,5	3,8	3,8	6,3	6,3	3,8
	Irat Var.	4,2	3,6	3,9	2,6	2,7	6,3
	Irat SD	2,1	1,9	2	1,6	1,6	2,5
GROUP 9	**Trat**	**9,4**	**9,4**	**10**	**10**	**10**	**10**
	IratMean	5,8	5,8	9,1	8,8	8,4	8,1
	Irat Max	8,8	7,5	10	10	10	10
	Irat Min	2,5	3,8	7,5	7,5	3,8	7,5
	IratVar.	5,2	3,5	1,4	2,1	9,8	1,6
	Irat SD	2,3	1,9	1,2	1,4	3,1	1,3
GROUP 10	**Trat**	**8,8**	**6,9**	**9,7**	**9,7**	**10**	**9,7**
	IratMean	4	5	5,8	6,8	8,1	7,8
	Irat Max	6,3	8,8	7,5	8,8	8,8	8,8
	Irat Min	1,3	1,3	3,8	5	6,3	7,5
	Irat Var.	4,2	9,4	2	2	1,6	0,3
	Irat SD	2,1	3,1	1,4	1,4	1,3	0,6
GROUP 11	**Trat**	**8,8**	**7,5**	**10**	**10**	**10**	**10**
	IratMean	4,8	5,8	7	6,9	8,3	5,5
	Irat Max	6,3	6,3	8,8	10	10	7,5
	Irat Min	2,5	5	2,5	1,3	6,3	1,3
	Irat Var.	4,2	0,5	6,7	15,1	2,8	6,7
	Irat SD	2,1	0,7	2,6	3,9	1,7	2,6
GROUP 12	**Trat**	**9,7**	**9,7**	**10**	**9,7**	**10**	**10**
	IratMean	6	7,5	6,5	8	9,5	8,4
	Irat Max	8,8	10	10	10	10	10
	Irat Min	2,5	5	2,5	5	7,5	6,3
	Irat Var.	5,9	3,9	7,1	3,6	1,3	2,5
	Irat SD	2,4	2	2,7	1,9	1,1	1,6
GROUP 13	**Trat**	**10**	**9,7**	**9,7**	**10**	**10**	**9,7**
	IratMean	7,2	7,5	7,5	9,1	9,1	8,1
	Irat Max	10	7,5	10	10	10	8,8
	Irat Min	2,5	7,5	3,8	8,8	7,5	7,5
	Irat Var.	10,8	0	7,3	0,4	1,4	0,5
	Irat SD	3,3	0	2,7	0,6	1,2	0,7
GROUP 14	**Trat**	**6,9**	**10**	**10**	**10**	**10**	**9,4**
	Irat Mean	2,5	6,3	8,1	9,2	9,5	6,7
	Irat Max	3,8	7,5	10	10	10	8,8
	Irat Min	1,3	5	5	7,5	8,8	2,5
	Irat Var.	1,6	1,9	4,2	1,7	0,5	4,8
	Irat SD	1,3	1,4	2,1	1,3	0,7	2,2

Continued in the next page…

	Stats	TBL1	TBL2	TBL3	TBL4	TBL5	TBL6
GROUP 15	**Trat**	**8,4**	**9,7**	**10**	**10**	**10**	**10**
	Irat Mean	4,4	7,5	6,3	7,7	9,1	7,9
	Irat Max	6,3	8,8	8,8	8,8	10	10
	Irat Min	1,3	5	3,8	6,3	7,5	5
	Irat Var.	3	1,9	3,1	1,5	1,4	2,9
	Irat SD	1,7	1,4	1,8	1,2	1,2	1,7
GROUP 16	**Trat**	**8,8**	**9,4**	**10**	**10**	**10**	**10**
	Irat Mean	4,8	6,5	7,3	8,5	7,9	8
	Irat Max	8,8	7,5	10	8,8	10	10
	Irat Min	1,3	3,8	5	7,5	3,8	1,3
	Irat Var.	7,8	2,7	4	0,3	7,9	14,5
	Irat SD	2,8	1,6	2	0,6	2,8	3,8
GROUP 17	**Trat**	**9,7**	**9,1**	**10**	**10**	**10**	**9,1**
	Irat Mean	6,3	5	6,3	7,3	9,4	7
	Irat Max	7,5	6,3	8,8	10	10	8,8
	Irat Min	5	2,5	2,5	2,5	8,8	5
	Irat Var.	0,8	3,1	7	8,1	0,8	2
	Irat SD	0,9	1,8	2,7	2,9	0,9	1,4
GROUP 18	**Trat**	**10**	**10**	**10**	**10**	**10**	**10**
	Irat Mean	6,5	8,3	9,5	8,8	9,1	9,3
	Irat Max	10	10	10	10	10	10
	Irat Min	3,8	5	7,5	7,5	6,3	8,8
	Irat Var.	5,3	4,4	1,3	0,8	3,5	0,5
	Irat SD	2,3	2,1	1,1	0,9	1,9	0,7
GROUP 19	**Trat**	**7,5**	**7,5**	**10**	**8,8**	**10**	**10**
	Irat Mean	5	5,8	5,2	7,5	8,3	7,3
	Irat Max	8,8	6,3	8,8	8,8	10	10
	Irat Min	2,5	5	0	6,3	2,5	3,8
	Irat Var.	6,9	0,5	9,6	1	10,6	5,8
	Irat SD	2,6	0,7	3,1	1	3,3	2,4
GROUP 20	**Trat**	**9,4**	**9,4**	**10**	**9,4**	**10**	**9,4**
	Irat Mean	5,5	8,4	9,4	6,9	10	7,1
	Irat Max	7,5	10	10	7,5	10	7,5
	Irat Min	1,3	7,5	8,8	6,3	10	6,3
	Irat Var.	6,7	1,4	0,5	0,5	0	0,5
	Irat SD	2,6	1,2	0,7	0,7	0	0,7
GROUP 21	**Trat**	**9,1**	**9,4**	**9,7**	**10**	**10**	**10**
	Irat Mean	5,2	5	7	6,7	6,5	5,6
	Irat Max	7,5	8,8	8,8	10	8,8	6,3
	Irat Min	3,8	2,5	5	3,8	1,3	3,8
	Irat Var.	2,1	6,9	2,8	4,8	9,7	1,6
	Irat SD	1,5	2,6	1,7	2,2	3,1	1,3

	Stats	TBL1	TBL2	TBL3	TBL4	TBL5	TBL6
GROUP 22	**Trat**	**9,4**	**8,1**	**10**	**9,7**	**10**	**10**
	Irat Mean	5,8	5,4	8,8	8,1	9,2	8,5
	Irat Max	8,8	7,5	10	10	10	10
	Irat Min	3,8	2,5	7,5	6,3	6,3	7,5
	Irat Var.	2,9	2,9	0,6	2,3	2,3	1,1
	Irat SD	1,7	1,7	0,8	1,5	1,5	1
GROUP 23	**Trat**	**10**	**8,4**	**9,7**	**10**	**9,7**	**9,7**
	Irat Mean	4,8	5	6,7	8,3	8,3	7,3
	Irat Max	7,5	7,5	8,8	10	10	8,8
	Irat Min	2,5	1,3	3,8	7,5	7,5	6,3
	Irat Var.	4,6	6,3	3,5	1,3	1	1,5
	Irat SD	2,2	2,5	1,9	1,1	1	1,2
GROUP 24	**Trat**	**9,4**	**8,8**	**10**	**9,7**	**10**	**9,1**
	Irat Mean	6	6	9,4	7,3	9,5	6,9
	Irat Max	8,8	7,5	10	8,8	10	7,5
	Irat Min	3,8	5	8,8	6,3	8,8	3,8
	Irat Var.	3,4	0,9	0,5	0,9	0,5	2,3
	Irat SD	1,8	0,9	0,7	0,9	0,7	1,5
GROUP 25	**Trat**	**8,8**	**7,5**	**10**	**9,7**	**10**	**9,4**
	Irat Mean	6,8	6,5	9,4	7,8	10	6,6
	Irat Max	8,8	7,5	10	8,8	10	8,8
	Irat Min	5	5	7,5	7,5	10	3,8
	Irat Var.	2	1,1	1,6	0,4	0	4,6
	Irat SD	1,4	1	1,3	0,6	0	2,1
GROUP 26	**Trat**	**7,5**	**8,8**	**9,7**	**9,7**	**10**	**9,4**
	Irat Mean	4,8	6	6,3	8,3	9,2	7,9
	Irat Max	6,3	7,5	7,5	10	10	8,8
	Irat Min	1,3	3,8	2,5	6,3	6,3	6,3
	Irat Var.	3,4	1,5	3,8	1,7	2,3	1
	Irat SD	1,8	1,2	1,9	1,3	1,5	1
GROUP 27	**Trat**	**9,4**	**9,6**	**9,9**	**9,8**	**9,8**	**9,8**
	Irat Mean	7,9	6,3	7	7,8	7,2	8,4
	Irat Max	8,8	7,5	10	8,8	8,8	9,4
	Irat Min	6,3	5	2,5	7,5	5	6,3
	Irat Var.	2,1	1	8,3	0,4	3,5	2,2
	Irat SD	1,4	1	2,9	0,6	1,9	1,5

Notes: *Irat Mean* corresponds to the average score of the group members performing the exercise individually; *Irat Max* corresponds to the score of the best performer in the group; *Irat Min* corresponds to the score of the worst performer in the group; *Trat* is the score obtained by the group working together

Appendix A2 – Group gains

		TB L1	TB L2	TB L3	TB L4	TB L5	TB L6
G. 1	G. A. I. P.	2,8	2,4	1,7	1,5	0,8	2
	Best performer	0,9	0,6	0	0	0	0,9
	Worst performer	4,6	3,1	6,2	2,5	2,5	3,4
G. 2	G. A. I. P.	2,9	2,8	1,7	1,7	0	1,2
	Best performer	1,2	0	0	0	0	0
	Worst performer	5	7,5	5	7,5	0	2,5
G. 3	G. A. I. P.	2,8	3,2	2,5	1,9	0,8	1,3
	Best performer	1,3	1,2	0	1,2	0	0
	Worst performer	5	5	7,5	3,7	2,5	2,5
G. 4	G. A. I. P.	3,8	3,7	1,6	1,2	0	0,6
	Best performer	1,9	1,2	0	0	0	0
	Worst performer	6,9	7,5	3,7	2,5	0	1,2
G. 5	G. A. I. P.	3,8	1,3	3,5	2,6	0	2,2
	Best performer	2,8	0	0	1,9	0	0,9
	Worst performer	4,1	2,5	6,2	5,6	0	4,7
G. 6	G. A. I. P.	3,3	4,6	4,4	2	1,7	2,7
	Best performer	1,8	1,9	0,6	0,3	0	0,6
	Worst performer	4,3	8,1	9,4	4,1	7,5	5,6
G. 7	G. A. I. P.	4,6	4,2	2,1	1,6	1	3,4
	Best performer	2,5	2,5	0	0,9	0	0,9
	Worst performer	7,5	6,2	5	3,4	2,5	5,9
G. 8	G. A. I. P.	4	3,2	3,7	1,9	1,5	1,9
	Best performer	1,3	1,2	1,2	0	0	-0,6
	Worst performer	6,3	6,2	6,2	3,7	3,7	5,6

		TB L1	TB L2	TB L3	TB L4	TB L5	TB L6
G. 9	G. A. I. P.	3,6	3,6	0,9	1,2	1,6	1,9
	Best performer	0,6	1,9	0	0	0	0
	Worst performer	6,9	5,6	2,5	2,5	6,2	2,5
G. 10	G. A. I. P.	4,8	1,9	3,9	2,9	1,9	1,9
	Best performer	2,5	-1,9	2,2	0,9	1,2	0,9
	Worst performer	7,5	5,6	5,9	4,7	3,7	2,2
G. 11	G. A. I. P.	4	1,7	3	3,1	1,7	4,5
	Best performer	2,5	1,2	1,2	0	0	2,5
	Worst performer	6,3	2,5	7,5	8,7	3,7	8,7
G. 12	G. A. I. P.	3,7	2,2	3,5	1,7	0,5	1,6
	Best performer	0,9	-0,3	0	-0,3	0	0
	Worst performer	7,2	4,7	7,5	4,7	2,5	3,7
G. 13	G. A. I. P.	2,8	2,2	2,2	0,9	0,9	1,6
	Best performer	0	2,2	-0,3	0	0	0,9
	Worst performer	7,5	2,2	5,9	1,2	2,5	2,2
G. 14	G. A. I. P.	4,4	3,7	1,9	0,8	0,5	2,7
	Best performer	3,1	2,5	0	0	0	0,6
	Worst performer	5,6	5	5	2,5	1,2	6,9
G. 15	G. A. I. P.	4	2,2	3,7	2,3	0,9	2,1
	Best performer	2,1	0,9	1,2	1,2	0	0
	Worst performer	7,1	4,7	6,2	3,7	2,5	5
G. 16	G. A. I. P.	4	2,9	2,7	1,5	2,1	2
	Best performer	0	1,9	0	1,2	0	0
	Worst performer	7,5	5,6	5	2,5	6,2	8,7

Continued in the next page…

		TB L1	TB L2	TB L3	TB L4	TB L5	TB L6
G. 17	G. A. I. P.	3,4	4,1	3,7	2,7	0,6	2,1
	Best performer	2,2	2,8	1,2	0	0	0,3
	Worst performer	4,7	6,6	7,5	7,5	1,2	4,1
G. 18	G. A. I. P.	3,5	1,7	0,5	1,2	0,9	0,7
	Best performer	0	0	0	0	0	0
	Worst performer	6,2	5	2,5	2,5	3,7	1,2
G. 19	G. A. I. P.	2,5	1,7	4,8	1,3	1,7	2,7
	Best performer	-1,3	1,2	1,2	0	0	0
	Worst performer	5	2,5	10	2,5	7,5	6,2
G. 20	G. A. I. P.	3,9	1	0,6	2,5	0	2,3
	Best performer	1,9	-0,6	0	1,9	0	1,9
	Worst performer	8,1	1,9	1,2	3,1	0	3,1
G. 21	G. A. I. P.	3,9	4,4	2,7	3,3	3,5	4,4
	Best performer	1,6	0,6	0,9	0	1,2	3,7
	Worst performer	5,3	6,9	4,7	6,2	8,7	6,2

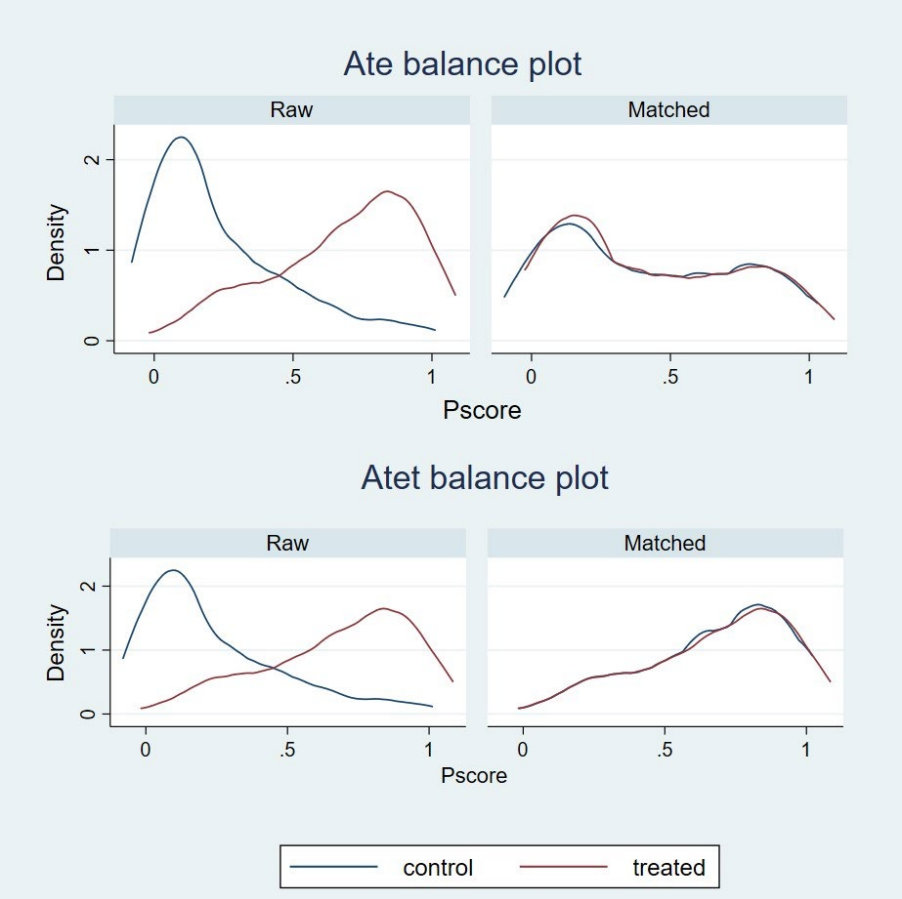
		TB L1	TB L2	TB L3	TB L4	TB L5	TB L6
G. 22	G. A. I. P.	3,6	2,7	1,2	1,6	0,8	1,5
	Best performer	0,6	0,6	0	-0,3	0	0
	Worst performer	5,6	5,6	2,5	3,4	3,7	2,5
G. 23	G. A. I. P.	5,2	3,4	3	1,7	1,4	2,4
	Best performer	2,5	0,9	0,9	0	-0,3	0,9
	Worst performer	7,5	7,1	5,9	2,5	2,2	3,4
G. 24	G. A. I. P.	3,4	2,8	0,6	2,4	0,5	2,2
	Best performer	0,6	1,3	0	0,9	0	1,6
	Worst performer	5,6	3,8	1,2	3,4	1,2	5,3
G. 25	G. A. I. P.	2	1	0,6	1,9	0	2,8
	Best performer	0	0	0	0,9	0	0,6
	Worst performer	3,8	2,5	2,5	2,2	0	5,6
G. 26	G. A. I. P.	2,7	2,8	3,4	1,4	0,8	1,5
	Best performer	1,2	1,3	2,2	-0,3	0	0,6
	Worst performer	6,2	5	7,2	3,4	3,7	3,1
G. 27	G. A. I. P.	1,5	3,3	2,9	2	2,6	1,4
	Best performer	0,6	2,1	-0,1	1	1	0,4
	Worst performer	3,1	4,6	7,4	2,3	4,8	3,5

Notes:
Cells report the gains of working in group. These are measured through the distance of the Trat score from:

d) The Group average individual performance (*G. A. I. P.*) ($X_{TRATg} - \overline{x_{ig}}$) ;
e) The Best performer ($TRat_g - \max_{Irat_i \in G_g}(IRat_i)$;
f) The Worst performer. ($TRat_g - \min_{Irat_i \in G_g}(IRat_i)$.

with i (1,…. 6) ; g (1, 27)

Appendix A3 – Kernel density plot of the propensity score matching



Appendix A4 – Covariance Balance summary

	Raw	Matched Panel A	Matched Panel B	Weighted Panel C
N° of obs	205	410	174	205
Treated obs	87	205	87	108
Control obs	118	205	87	97

	Standardized differences		Variance ratio	
	(a) Raw	(b) Weighted	(a) Raw	(b) Weighted
PANEL A - propensity-score matching ATE				
Female	-0.57	0.12	1.02	0.98
Fuoricorso	-0.01	-0.11	0.97	0.63
Worker	-0.31	-0.13	0.68	0.81
Fuorisede	-0.27	0.43	0.55	1.78
Working experience	0.07	0.05	0.89	0.92
High school[base L_Class]				
L_Scie	0.49	0.01	1.27	1.01
L_Ling	-0.60	-0.28	0.27	0.56
L_Other	-0.07	-0.06	0.69	0.67
Technical	-0.01	0.17	1.00	1.07
Professional	0.03	0	1.36	1
Microscore[base pending exam]				
18to22	-0.36	-0.15	0.51	0.79
23to26	-0.23	-0.20	0.81	0.78

27to30	0.84	0.06	1.63	1.04
Team	0.31	0.12	0.79	0.88
Extyrovert	0.02	0.05	1.00	0.99
Leader	-0.41	-0.10	1.05	1.04
Advocancy_Role	-0.15	-0.40	0.81	0.57
Listening_Role	0.19	0.07	1.29	1.12
EcoSkill	-0.60	-0.30	0.76	0.79
EcoPass	-0.19	-0.13	1.02	0.95
MathSkill	0.51	0.23	1.65	1.40
MathPass	-0.07	0	0.81	1
PANEL B - propensity-score matching ATET				
Female	-0.57	0.05	1.02	1.03
Fuoricorso	-0.01	-0.05	0.97	0.84
Worker	-0.31	0.28	0.68	1.93
Fuorisede	-0.27	0.04	0.55	1.13
Working experience	0.07	0.06	0.89	0.90
High school[base L_Class]				
L_Scie	0.49	0.02	1.27	1
L_Ling	-0.60	-0.13	0.27	0.65
L_Other	-0.07	0	0.69	1
Technical	-0.01	-0.05	1.00	0.97
Professional	0.03	0.15	1.36	.
Microscore[base pending exam]				
18to22	-0.36	-0.29	0.51	0.57
23to26	-0.23	0.17	0.81	1.28
27to30	0.84	0.02	1.63	0.99
Team	0.31	-0.26	0.79	1.49
Extyrovert	0.02	-0.30	1.00	1.09
Leader	-0.41	-0.35	1.05	1.02
Advocancy_Role	-0.15	-0.47	0.81	0.63
Listening_Role	0.19	0.39	1.29	1.91
EcoSkill	-0.60	0.17	0.76	1.26
EcoPass	-0.19	0.46	1.02	1.32
MathSkill	0.51	0.62	1.65	2.02
MathPass	-0.07	0.20	0.81	2.22
PANEL C – Treatment probit model for the augmented IPW				
Female	-0.57	0.02	1.02	0.99
L_Scie	0.49	0.01	1.27	1.01
L_Class	-0.04	0.01	0.86	1.02
Microscore[base pending exam]				
18to22	-0.36	-0.00	0.51	0.99
23to26	-0.23	-0.09	0.81	0.92
27to30	0.84	0.01	1.63	1.01
MathSkill	0.51	0.04	1.65	1.06

Notes: for the Aipw estimates (Panel C) -as for description in the methodological section- all covariates enter in the equation of the treatment effect, and a reduced form (selection data driven on main determinants on treatment status) enter in the probit model which predict treatment condition

Acknowledgments

Funding from "Innovative and Inclusive Academia", FAR 2021 University of Modena and Reggio Emilia FOMO Line – Mission Oriented Supply Chain Interdisciplinary Research Project is gratefully acknowledged. We are grateful to the participants at the "Innovative and Inclusive Academia" International Final Conference held in the Department of Economics Marco Biagi, University of Modena and Reggio Emilia on 19th and 20th October, 2023 for their stimulating comments on a previous version of the paper. Responsibility for errors and omissions rests with the authors.

7. References

Ahn, M.Y. and Davis, H.H. (2020) 'Students' sense of belonging and their socio-economic status in higher education: a quantitative approach', *TEACHING IN HIGHER EDUCATION* [Preprint]. doi:10.1080/13562517.2020.1778664.

Allgood, S. and McGoldrick, K. (2021) 'Team-Based Learning in Economics: A Symposium', *Journal of Economic Education*, 52(3), pp. 218–219. doi:10.1080/00220485.2021.1925190.

AlmaLaurea (2023) *XXV Indagine Profilo dei Laureati 2022. Rapporto 2023*, https://www.almalaurea.it/

Cagliesi G, Ghanei M (2022) Team-based learning in economics: Promoting group collaboration, diversity and inclusion. *The Journal of Economic Education*, 53:11–30. https://doi.org/10.1080/00220485.2021.2004276.

Carson KS, Adams H, Gonzalez-Ramirez J, et al (2021) Challenges and lessons: Design and implementation of a multi-site evaluation of team-based learning. *The Journal of Economic Education* 52:241–248. https://doi.org/10.1080/00220485.2021.1925185 .

Clerici-Arias M (2021) Transitioning to a team-based learning principles course. *The Journal of Economic Education* 52:249–256. https://doi.org/10.1080/00220485.2021.1925184 .

Espey M (2022) Variation in individual engagement in team-based learning and final exam performance. *International Review of Economics Education* 41:100251. https://doi.org/10.1016/j.iree.2022.100251

European Commission (2020) Commission launches Youth Employment Support: a bridge to jobs for the next generation. https://ec.europa.eu/social/main.jsp?catId=89%26newsId=9719%26furtherNews=yes%26langId=en

Good C, Rattan A, Dweck CS (2012) Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of Personality and Social Psychology* 102:700–717. https://doi.org/10.1037/a0026659

Goodenow, C. (1993) "Classroom Belonging among Early Adolescent Students." *The Journal of Early Adolescence* 13 (1): 21–43.

Imai, K., and M. Ratkovic (2014) Covariate balancing propensity score. *Journal of the Royal Statistical Society*, Series B 76: 243–263. https://doi.org/10.1111/rssb.12027

Kane, S., Chalcraft, D. and Volpe, G. (2014) 'Notions of belonging: First year, first semester higher education students enrolled on business or economics degree programmes', *International Journal of Management Education*, 12(2), pp. 193–201. doi:10.1016/j.ijme.2014.04.001.

Koutsouris, G., Stentiford, L., & Norwich, B. (2022) A critical exploration of inclusion policies of elite UK universities. *British Educational Research Journal*, 48(5), 878-895–895. https://doi.org/10.1002/berj.3799

Lorenzo-Lledó, A., Lorenzo Lledó, G., Lledó, A., & Pérez-Vázquez, E. (2023). Inclusive education at university: a scientific mapping analysis. *Quality and Quantity*. https://doi.org/10.1007/s11135-023-01712-w

Michaelsen, L. K., Knight, A. B., & Fink, L. D. (2004) Team-based learning: A transformative use of small groups in higher education. Sterling, VA: Stylus.

Mann, H. B., and D. R. Whitney (1947) On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics* 18: 50–60.

Moore, VJ, Prewitt, EM, Carpenter-McCullough, AJ & Whitworth, BA (2020) 'Teamwork Makes the Dream Work: Using Team-Based Learning in the Science Classroom', *Journal of College Science Teaching*, vol. 49, no. 3, p. 38

Moriña A. (2017) Inclusive education in higher education: challenges and opportunities, *European Journal of Special Needs Education*, 32:1, 3-17, https://doi.org/10.1080/08856257.2016.1254964.

Moriña, A., Orozco, I. (2022) "Inclusive learning strategies at university: the perspective of Spanish faculty members from different knowledge areas". *Cult. Educ.* 34(2), 231–265.

OECD (2023), *Education at a Glance 2023: OECD Indicators*, OECD Publishing, Paris, https://doi.org/10.1787/e13bef63-en.

Opdecam E, Everaert P, Van Keer H, Buysschaert F (2013) Preferences for Team Learning and Lecture-Based Learning Among First-Year Undergraduate Accounting Students. *Res High Educ.* https://doi.org/10.1007/s11162-013-9315-6

Parmelee DX, DeStephen D, Borges NJ (2009) Medical Students' Attitudes about Team-Based Learning in a Pre-Clinical Curriculum. *Medical Education Online* 14:4503. https://doi.org/10.3402/meo.v14i.4503

Rubin, D. B. (1973) Matching to remove bias in observational studies. *Biometrics* 29: 159–183. https://doi.org/10.2307/2529684.

Ruder, P., M. H. Maier, and S. P. Simkins (2021) Getting started with TBL: An introduction. *Journal of Economic Education*, 52 (3): 220–30.

Schell J. A., Butler A. C. (2018). Insights from the science of learning can inform evidence-based implementation of Peer instruction. *Frontiers in Education*, 3, 33. 10.3389/feduc.2018.00033

Schwartz, D. L., Tsang, J. M., & Blair, K. P. (2016) *The ABCs of how we learn: 26 scientifically proven approaches, how they work, and when to use them*. W W Norton & Co.

Simkins, S.P., Maier, M.H. and Ruder, P. (2021) 'Team-Based Learning (TBL): Putting Learning Sciences Research to Work in the Economics Classroom', *Journal of Economic Education*, 52(3), pp. 231–240. doi:10.1080/00220485.2021.1925188.

Sparrow, Malcolm K. (2011) "The GRumbler". Latest update: 11th October 2021 https://scholar.harvard.edu/msparrow/grumbler

StataCorp (2021) 'Stata treatment effect reference manual: potential outcomes/counterfactual outcomes' *Stata: Release 17. Statistical Software*. College Station, TX: StataCorp LLC. ISBN-13: 978-1-59718-351-2.

Thomas, L. (2012) *Building Student Engagement and Belonging in Higher Education at a Time of Change*. Paul Hamlyn Foundation. https://www.phf.org.uk/wp-content/uploads/2014/10/What-Works-reportfinal.pdf

Tremblay-Wragg, E., Raby, C., Ménard, L., Plante, I. (2019) The use of diversified teaching strategies by four university teachers: What contribution to their students' learning motivation? *Teach. High. Educ.* 26(1), 97–114. https:// doi. org/ 10. 1080/ 13562 517. 2019. 16362 21

United Nations (2015) *Transforming our world: the 2030 agenda for sustainable development*. A/RES/70/1. https:// sdgs.un.org/ 2030agenda

Watson, W. E., Michaelsen, L. K., & Sharp, W. (1991) Member competence, group interaction, and group decision making: A longitudinal study. *Journal of Applied Psychology*, 76(6), 803–809. https://doi.org/10.1037/0021-9010.76.6.803.

Weiss DJ, Dawis RV, England GW, Lofquist LH. (1967) *Manual for the Minnesota Satisfaction Questionnaire*. Minneapolis (MN): University of Minnesota.

Wilcoxon, F. (1945) Individual comparisons by ranking methods. *Biometrics* 1: 80–83.