Gruppo di lavoro sulle tecniche
di insegnamento e di apprendimento

https://riviste.unige.it/index.php/glia/index
2975-0075

N°4- Anno 2025
pp. 1-35

TEAM-BASED LEARNING AND GENDER INCLUSIVENESS IN ECONOMICS

Tindara ADDABBO[1], Chiara TASSELLI[2], Filippo DAMIANI[3]

1 Dipartimento di Economia Marco Biagi, Modena (MO), tindara.addabbo@unimore.it
2 Dipartimento di Economia Marco Biagi, Modena (MO), chiara.tasselli@unimore.it
3 IDEM, Modena (MO), f.damiani@idemindthegap.it

GENOVA UNIVERSITY PRESS

Abstract

Women's low self-confidence in quantitative fields affects both academic achievement and career choices, perpetuating gender inequality. This structural problem contributes to horizontal segregation, hinders progress toward SDG 5 and results in a substantial loss of talent in the economic system. This paper investigates the impact of implementing team-based learning (TBL) on student performance in macroeconomics at a university in Northern Italy. Utilizing a structural break in the academic year when TBL was introduced, the study assesses its effectiveness and explores potential gender biases in outcomes. The econometric analysis employs multivariate regression, probit analysis, and a Cragg model to measure TBL's effect on grades while controlling for student socio-demographic characteristics. Results indicate a positive impact on student performance. Particularly females, demonstrate improved macroeconomic scores, meanwhile, males experience a significant increase in the likelihood of passing the exam, indicating shifts not only in performance but also in attitudes and exam approach.

Keywords

Gender economics, Team-Based Learning, Gender bias, Macroeconomics education, Academic performance

1. Introduction

Previous studies in the education field have shown that gender plays a decisive role in students' academic enrolment and performance on university courses in economics, science and technology. Females' low self-confidence in quantitative and scientific courses is found to shape both their performances (Ballard and Johnson 2008) their educational choices (Di Tommaso, Macagnan and Mendolia 2021) and therefore their access to employment and horizontal segregation in the labour market (Borrowman and Stephan 2020). That structural problem - in addition to perpetuating gender inequality in this field and making it difficult to reach SDG 5 - results in a massive loss of talent in the economic system. This paper focuses on the impact of a change in teaching methodology, from traditional teacher-centred lecturing to a methodology that has been proven to promote active learning and teamwork, namely Team-Based Learning (TBL). TBL was developed by Michaelsen in the late 1970s and has been increasingly used in the US since the 1980s in a variety of disciplines in tertiary education, though its application to economics has been limited (Cagliesi and Ghanei, 2022). TBL has been introduced in an undergraduate course in Macroeconomics within a wider project carried out by a public university in the North of Italy, based on its expected positive impact on students' soft skills in problem-solving and teamwork development, within a wider project on developing soft skills in that university. As assessed by Simkins, Maier, and Ruder (2021), TBL intentionally promotes learning strategies that learning sciences research has identified as highly effective in creating powerful learning environments for students. The attention paid to the group's composition resulting in within-group diversity, also as regards gender, also allowed us to test its impact on inclusion. The analysis compared students' performance through a consistent and robust estimator for models with censored data, using Cragg's model (or two-part, Hurdle model) on a sample of 711 students attending a macroeconomics course in a bachelor course in a large public university in North-Eastern Italy.

It can, therefore, be hypothesized that:

H1 = Attending Team-Based Learning Lessons produces better learning outcomes.

H2 = Students react differently to the treatment depending on their gender.

H3 = Female performance in Macroeconomics is lower.

H4 = TBL could help overcome gender differences in macroeconomics.

We do try to reply to these research questions or at least try to get a clearer view of the relationship between students, their progression in the learning process and macroeconomics classes (also from a gender perspective).

2. TEAM-BASED LEARNING METHODOLOGY AND EXPECTED OUTCOMES

The focus of this paper is on the impact of a particular methodology that has been recognised in the literature as able to develop students' active engagement and specific, soft skills and, in its implementation, allows a high degree of inclusiveness: Team-based learning (TBL). TBL was developed by Michaelsen in the late 1970s, and has been increasingly used in the US since the 1980s in a variety of disciplines in tertiary education though its application to economics has been limited (Cagliesi and Ghanei 2022). Michaelsen et al. (2004a), described TBL as an unusually powerful and versatile teaching strategy that enables teachers to take small-group learnings to a new level of effectiveness. TBL group work has been found to be powerful in improving the ability of students to apply course contents since, during TBL activities, the development of self-managed learning teams is promoted (Michaelsen and Sweet, 2008; Michaelsen, Davidson and Major 2014).

The composition of TBL teams and their duration play a crucial role in the efficacy of the approach and its inclusive content. TBL teams are formed and the membership of the groups must be kept stable during the whole term to allow team development (Michaelsen, Watson and Sharp 1991). Care must be taken on the composition of the groups since it has indeed been demonstrated that the most effective results are obtained in groups with the most diverse composition (Phillips et al. 2008; Parmelee and Michaelsen, 2010), which means that groups are deliberately formed to be diverse and cohesive (Kathleen and Odell 2018). The dimension of groups is five to seven members in order to ensure the group dimension that is considered efficient to face the variety of decision-based tasks encountered during TBL implementation (Michaelsen et al. 2004b).

TBL can be considered a student-centred class methodology. Students are assigned course materials before a teaching session (flipped classroom) to be able to apply in classes their self-gained knowledge (Balan et al. 2015). In-class activities are typically based on the Readiness Assurance Process (RAP), which consists of two Readiness Assurance Tests (RAT) in which the students should answer the same questions first individually (iRAT, Step one), and then as a team (tRAT, Step two). Then, after the instructor's clarification lecture on the first set of questions, students work again on a team application (tAPP, Step 3). As stated by Espey (2018):

> Significant problems engage students in concrete examples so they understand the usefulness of the course concepts. Specific choices require teams to take a position, sometimes also requiring them to support that position with a short rationale of their choice. Forcing all students to confront the same problem enables them to better engage with each other across teams, while simultaneous reporting precludes teams from simply agreeing with the majority of others, forcing them to decide before knowing what other groups will say. [Espey, 2018, p.10]

The fourth part of the activities consists of peer assessment and feedback, leading to students' evaluation of their team-mates (Step 4); this last part is fundamental to enhance the ability to work together and positively contribute to the team (Michaelsen, Davidson and Major 2014) and to avoid free-riding (Hettler 2015). While frequently implemented in a face-to-face classroom, TBL has received limited attention in the online learning environment where geographically distributed, and asynchronous learning poses challenges to its fundamental design (Goh et al. 2020). Virtual reality could be a platform to provide the engaging elements of TBL, without students needing to be physically present in the same room. It has the potential to be a useful tool for online, distance TBL (Coyne et al. 2018).

Among the positive impacts of TBL, the literature has shown increased student engagement both in class and out of class (Espey, 2012; Imazeki 2015; Ruder, Maier and Simkins 2021) and increased attendance (Abio et al. 2019). Evidence has been provided on the positive impact of the adoption of TBL in show-up percentages of students at the final exam and in their rate of success in passing the exam for students re-taking a subject (Abio et al. 2019) and for students in STEMM courses (Parappilly et al., 2021). Evaluation of the implementation of TBL in courses on the principles of microeconomics and quantitative methods as compared to lecture-based instruction, allowed Hettler (2015) to detect differences in the outcome of TBL on the exam scores for different groups of students namely, the minority and first-generation college students' status show a positive and significant marginal impact on exam scores in TBL sections thus supporting the hypothesis that TBL can have a greater impact on groups that are typically disadvantaged. Cagliesi and Ghanei (2022) found evidence of a positive impact of TBL on grades in economic courses and a reduction in the attainment gap for Black, Asian, and minority ethnic students. In terms of the efficacy of the TBL methodology in fostering inclusion, evidence has been provided of the reduction in achievement gaps for minorities attending courses using TBL sessions, as well as evidence on the TBL approach being more attractive for female and non-white students (Clerici-Arias, 2021).

Another line of investigation in the evaluation of TBL concerns the impact of the teams' characteristics on team or individual outcomes or of the behaviour in teams on individual outcomes. Espey (2018) analyzes what measurable team characteristics influence team and individual performance in the comprehensive final exam. The latter has been found to be positively affected both for men and for women by a more equal gender distribution within TBL groups. Espey (2022) shows evidence of a positive impact on final exam scores of increased effort or engagements in team-based activities.

3. TEAM-BASED LEARNING IMPLEMENTATION AND THE GENERATED DATA

TBL methodologies had already been adopted in the University we have analyzed here, in 2017 within the project teaching for competencies involving about 2,000 students in the experimentation showing a positive impact on the development of soft skills considered fundamental in business contexts (De Santis et al, 2019, Bellini et al, 2020). Through contacts with stakeholders (companies, public and private bodies, the tertiary sector) the university analyzed identified the two soft skills that at the beginning of the project were the most requested by the labour market: problem-solving, i.e. an approach to work that, by identifying priorities and critical issues, allows those involved to identify the best possible solutions to problems; team-work, i.e. the willingness to work and collaborate with others, having the desire to build positive relationships aimed at achieving the assigned task. TBL was thus chosen as a methodology able to develop these soft skills and implemented in the academic year 2017/2018 in sixteen courses with sixteen control courses that allowed an evaluation of the impact of TBL on students' soft skills. Instructors and tutors involved in the TBL courses were involved in a training course to acquire knowledge on TBL methodology and how to restructure their syllabus. A community of practices was then built within the university that experimented the TBL method in strict collaboration with the Italian National TBL Community of practices.
The undergraduate course in Macroeconomics analyzed in this paper was involved from the very beginning of the TBL implementation and the data collected refer in total to 891 students (1,345 including those who resat their exam) who attended the course from the academic year 2016/2017 (when TBL had not yet been implemented) to the academic year 2020/2021. To ensure diversity, the groups were created using G(roup)Rumbler, an algorithm developed by Malcolm K. Sparrow in 2011 to ensure higher heterogeneity within groups of students across the class (Sparrow, 2011). The variables that were used in this implementation of the GRumbler to form the TBL groups were collected throughout a survey run before TBL classes in each academic year and refer to gender, age, origin, type of secondary school attended, grades in Maths and

Microeconomics, students' attitude in team working, personal characteristics, etc. The goal was to allow within-group diversity in line with what has been found to increase the effectiveness of TBL in developing teamwork and problem-solving and also to have a positive impact on inclusiveness. The group membership was kept permanent with semester-long teams. TBL was implemented in the Macroeconomics semester course, which is structured in thirty lectures, using active learning techniques and six TBL units with partial pre-class assignments following the Readiness Assurance Process four-step structure described in Section 2.

3.1 The Sample

Students were cohorts attending the second year of the Undergraduate Course in Macroeconomics from the academic year 2016/2017 till 2020/2021. To avoid any possible contamination of the data caused by the occurrence of the pandemic, we decided to cut the sample in February 2020. Before this date, both the teaching and examination methods remained practically unchanged, except for the introduction of the TBL in the academic year 2017/2018. The lectures took place in the first Semester of the academic year from September to December. A total of six exams are taken in each academic year: two in the Winter Session, three in the Summer Session and one in the Fall Session. The final sample, therefore, consists of 711 students and 1,024 exam attempts. Out of the 1,024 examination tests, 439 were carried out by female students and the remaining 585 by male students. The distribution between sessions tends to be concentrated in the winter session, the closest to the semester when the course is taught. Female students have an average attendance rate in TBL nearly five percentage points higher than that of men. A deeper insight concerning the distribution of the sample, its relation to the treatment and its respective characteristics will be addressed later.

3.2 Data and Variables description

To investigate the gender differences in macroeconomics among the undergraduate students analyzed, multiple data sources have been merged. Administrative data have been downloaded for the purpose from the student management system, including results of intermediate tests collected by the professor and socio-demographic and behavioural covariates obtained by getting students to respond to a questionnaire.

Measuring Variables

Dependent and independent variables are defined as follows (a more detailed description of the variables is provided in Table A1 in Appendix 1):

There are two dependent variables used to represent students' academic performance: a continuous variable which reports the students' final grade in macroeconomics (Mark) and a dummy variable (Pass) stating if the student passed or failed the exam.

It is important to stress that the selection of the Mark variable implies our sample being classified as censored from the below (and above) sample. The latter is representative of the population because all students who attempted the macroeconomics exam at least once are in the sample, but the mean of the dependent variable is not because we cannot observe students' marks if they fail the exam as we do not know their true performance if they succeeded. This means that the variable has a lower bound set on the score of 17, and for students who cannot reach it we cannot observe the actual performance. The same applies to the highest extremity of the distribution where there is an upper bound at 30 cum laude (we cannot observe the real mark over thirty). On the other hand, several explanatory variables are used in this study.

Some related to students' academic paths like:

i) The university entrance score at TOLC (TEST ONLINE CISIA) in Maths (MathAbility)

ii) The university entrance score at TOLC in Logic (LogicAbility)

iii) The university entrance score at TOLC in reading comprehension (ComprehensionAbility)

Points (i, ii and iii) are considered proxies of ability before entering university.

iv) if they have already attempted the exam (Retaker)

v) whether they attend TBL classes or not (TBL) and the v) TBL dosage (Dosage).

vi) The number of credits obtained in the first year (Credits)

vii) the average of all the exams taken by the students during their academic career subdivided into 3 macro-groups (whose disaggregation is detailed in Table A2 in section 7 – addendum):

a. Highlyquantitative

b. Slightlyquantitative

c. Nonquantitative

Other covariates relating to students' sociodemographic characteristics are also included, namely:

viii) gender (Female)

ix) (LowIncome) as a low family income could adversely affect school performance

x) (Native) Italian nationality

In addition, time-fixed effects account for all unobservable factors that are changing across sessions.

4. Empirical Analysis

4.1 Research Design and Implementation

The research design reflects the methodologies adopted in the Introductory macroeconomics course that is the object of this study. Before the academic year 2017-2018, the course was held mainly in a traditional lecture-based format and, thereafter the same instructor changed the structure of the course by adopting the TBL approach. Groups of 5-6 members were formed according to the GRumbler algorithm referred to in Section 3 of this paper, which considered socio-demographic characteristics such as gender, ethnicity, openness, and scholastic skills. Team membership was kept stable throughout the duration of the semester and students worked together to solve the T-Rat and the case study (T-App), meanwhile they faced the I-Rat and the team-mates' evaluation individually. For each year in which TBL was implemented, the intervention dosage consisted of 6 sessions – lasting an hour and a half each – distributed in the semester, the rest consisted of lectures and classes where active participation of students was required as in the instructor's style of lecturing. We have, therefore, used the structural break from a lecture-based format to a TBL-based format in the same course and with the same instructor to evaluate TBL impact on students' achievements controlling for a set of variables that also includes the students' cohort.

Panel A: Before the introduction of TBL (Academic Year 2016/2017)

	(1) FEMALE		(2) MALE	
	No TBL	TBL	No TBL	TBL
N	105	NA	136	NA
P(fail)	24,8%	NA	22%	NA
$\underline{Mark}\vert_{Pass}$	24,1	NA	24,5	NA

Panel B: Before the introduction of TBL (Academic Years 2017/2018; 2018/2019; 2019; 2020)

	(1) FEMALE		(2) MALE	
	No TBL	TBL	No TBL	TBL
N	99	235	155	294
P(fail)	51%	35%	54,8	29%
$\underline{Mark}\vert_{Pass}$	20,8	24,4	22,2	24,1
Participation rate	70%		65%	

Table 1 – Detail on treatment and ITT

Table 1 shows the distribution of the sample by gender year of the exam and education path covered. What stands out in the figure is that the raw average mark in macroeconomics (calculated on sufficient marks only) for females who had not experienced TBL is lower than the male counterpart. Vice versa, females who attended TBL succeeded in getting a higher mark compared to males. This finding is consistent with the literature surveyed in Section 2, showing a reduction of certain groups of students' gap in achievements when the TBL methodology is adopted.

Table 1 also reveals that female students show a higher participation rate in TBL than male students again a result consistent with the literature and in line with the inclusive scope of the methodology application.

4.2 Methodology

The statistical analysis was carried out using the econometric package of STATA and start with a regression analysis as dominant empirical tool used by economists (Miller and Rodgers 2008).

At first, the regression model in Eq. 1 was estimated to detect what were the most significant variables for our analysis and their best combination both for the goodness of fit and the coefficient strength (definition of the model). *$Mark_i$* the Macroeconomic performance obtained by students: 30 is the maximum grade a student can achieve and 18 is the minimum grade that a student can get. All fail marks (which are unobserved) are set at 17.

α_0 is the model intercept, β_1 is our coefficient of interest reflecting the TBL effect on macroeconomics grades, X_i regards a set of students' characteristics: entry capabilities, the average grade in exams and some socio-demographic information for the description of variables. Finally, γ_i are session-fixed effects. The same equation is repeated also for the male (eq. 1b) and female (eq. 1a) subsample. In all the equations, the subscript i corresponds to students.

Then, the probit model in Eq. 2 was implemented to find out the probability of passing the macroeconomics exam.

$$Pass\{0\ if\ student\ fail\ \ 1\ if\ student\ pass$$

$Eq.2: Pass_i = \phi(\alpha_0 + \beta_1 T_i + \beta_2 X_i + \beta_3 \gamma_i + \varepsilon_i)$ with *i* (1,…585)

$Eq.2a: Pass_{i|F} = \phi(\alpha_0 + \beta_1 T_i + \beta_2 X_i + \beta_3 \gamma_i + \varepsilon_i)$ with *i* (1,…243)

$$Eq.2b: Pass_{i|M} = \phi(\alpha_0 + \beta_1 T_i + \beta_2 X_i + \beta_3 \gamma_i + \varepsilon_i) \quad \text{with } i \ (1,\ldots 342)$$

$\phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. Meanwhile, for all other terms, the functional form on the right side of the equation follows the decryption of equation 1 and, as with the regression model, our coefficient of interest is β_1 which multiplies treatment. Also in this case the equation is reproduced for the male (eq. 2b) and female (eq. 2a) subsamples.

Finally, a Cragg's model (two-part, Hurdle model in equation 3, 3a, 3b) was estimated to obtain the best possible fit for this sample in which the dependent variable is censored from both the bottom and the top. This model is a modified version of the Tobit model (Tobin 1958) and is preferred to the latter following the likelihood log test in equation 4.

We face a censored sample since the mark is not detectable below (or above) a defined threshold. Conveniently, unlike the truncated sample, the censored sample is representative of the population because all observations are included; only the dependent variable suffers losses of information.

In our analysis, the marks of the students who pass the exam are detectable and range from 18 to 30, while we cannot observe the scores of the students who fail the exam (they may have obtained a mark of 2 or 17 – the only information we have is that they did not achieve a pass mark). The model has, therefore, a lower limit at 17. The same concept holds for students who achieved top marks in the exam: it is not possible to distinguish the different marks within excellence. For that reason, we set also 30 as the upper bound.

$$Mark_i^* = \{17 \quad if\ Mark_i^* < 18\ Mark_i \quad if\ 18 \leq Mark_i^*\ 30 \quad if\ Mark_i^* \geq 30$$
$$< 30$$

$$Eq.3: Pass_i = \phi(\alpha_0 + \beta_1 T_i + \beta_2 X_i + \beta_3 \gamma_i + \varepsilon_i) \quad \text{with } i \ (1,\ldots 585)$$

$Eq. 3a: Pass_{i|F} = \Phi(\alpha_0 + \beta_1 T_i + \beta_2 X_i + \beta_3 \gamma_i + \varepsilon_i)$ with *i* (1,…243)

$Eq. 3b: Pass_{i|M} = \Phi(\alpha_0 + \beta_1 T_i + \beta_2 X_i + \beta_3 \gamma_i + \varepsilon_i)$ with *i* (1,…342)

Equation 3 following Tobin's specification shows the processing of the dependent variable within the model. The actual value for is observed if the latent variable is between 18 and 30, meanwhile the lower limit is observed for the censored from below observations and the upper limit is observed for the censored from above observations.

The combination of covariates used to estimate this functional form is the same as in equation 1 with the distinction that the dependent variable $Mark_i^*$ is not a simple continuous variable but it is assumed to be the latent. The Tobit model was applied in two-step (rather than one), relaxing the assumption that the discrete event and the continuous event are the same, allowing different coefficients for the probability of passing the exam and for the continuous grade variable, once a passing grade has been achieved (Cragg's model - a Tobit variant).

The decision to opt for Cragg's model was taken following a dedicated test for the best fit as displayed in equation 5. We estimated separately tobit, probit, and truncated regression (Cragg's) models and derived their log-likelihoods to compute the following likelihood ratio statistic:

Eq. 4: $\lambda = 2 \cdot (LL_{probit} + LL_{truncreg} - LL_{tobit})$

The application of the formula is reported in addendum A4 in Appendix 1 where the chi-square test validates the best fit of Cragg's model. This condition is verified both for the main model and for its gender disaggregation.

4.3 Descriptive Statistics and findings

Table 2 shows descriptive statistics for the covariates characterising our sample. Panel A is dedicated to continuous variables, whereas panel B is dedicated to dichotomous ones. Both panels consist of two subsections that provide gender comparisons between the control (1) and the treated

(2) group. The typical self-selecting students as treated have a higher average in all subject types (high, middle and non-quantitative); they have a short time-gap between completing secondary school and enrolling at university, and earned more credits in the first year of the undergraduate course attended.

Panel A.1: Continuous Variable for controls

	(1) MALE		(2) FEMALE		(3) T-TEST $(\overline{x_m} - \overline{x_f})$	
	mean	sd	mean	sd	b	t
Mark	20.97	4.62	20.70	4.42	0.28	(0.67)
Highly quant.	23.36	2.86	23.50	3.10	-0.13	(0.47)
Slightly quant.	22.59	2.94	22.44	2.72	0.15	(0.51)
Non quant.	23.60	2.43	23.91	2.45	-0.31	(-1.34)
Tolc[a]	16.45	5.18	12.39	5.16	4.06***	(8.46)
Compreh.ability	5.85	1.88	4.74	2.09	1.11***	(5.27)
Math Ability	4.52	3.02	3.04	2.51	1.48***	(5.18)
Logicability	6.25	2.24	5.40	2.52	0.85***	(3.36)
Tmaxingl[B]	20.84	4.06	21.60	4.19	-0.76	(-0.85)
Dosage	0.05	0.46	0.15	0.63	-0.10	(-1.84)
Iratscore[C]	0.12	1.12	0.80	3.41	-0.68**	(-2.74)
Attempts	1.52	0.97	1.49	0.84	0.03	(0.37)
Gap_diploma	0.38	2.52	0.50	1.60	-0.12	(-0.66)
Credits	38.11	16.69	37.75	15.80	0.36	(0.24)

Panel A.2: Continuous Variable for treated

	(1) MALE		(2) FEMALE		(3) T-TEST $(\overline{x_m} - \overline{x_f})$	
	mean	sd	mean	sd	b	t
Mark	22.02	4.90	21.80	4.99	0.22	(0.50)
Highly quant.	24.73	3.04	24.77	3.25	-0.03	(-0.11)
Slightly quant.	24.21	2.97	23.56	3.07	0.65*	(2.00)
Non quant.	24.70	2.51	24.60	2.38	0.10	(0.44)
Tolc[a]	17.19	6.10	14.24	5.53	2.95***	(5.73)
Compreh.ability	5.56	2.40	5.06	2.22	0.50*	(2.26)
Math Ability	5.02	3.01	3.57	3.06	1.44***	(4.93)
Logicability	6.85	2.53	5.35	2.46	1.49***	(6.23)
Tmaxingl[B]	20.75	5.10	20.27	5.21	0.49	(0.96)
Dosage	5.82	0.39	5.80	0.40	0.02	(0.69)
Iratscore[C]	16.08	8.37	16.75	7.61	-0.67	(-0.96)
Attempts	1.34	0.63	1.44	0.73	-0.11	(-1.75)
Gap_diploma	0.21	0.73	0.24	0.79	-0.03	(-0.45)
Credits	39.91	16.06	35.62	15.99	4.29**	(3.02)

Panel B.1.: Dichotomous Variable for control

	(1) MALE		(2) FEMALE		(3) T-TEST $(\overline{x_m} - \overline{x_f})$	
	mean	sd	mean	sd	b	t
Pass	0.60	0.49	0.63	0.48	-0.02	(-0.51)
Retaker	0.32	0.47	0.32	0.47	-0.00	(-0.09)
cred40	0.51	0.50	0.46	0.50	0.05	(1.15)
native	0.96	0.20	0.85	0.36	0.11***	(3.80)

LowIcome	0.06	0.23	0.25	0.43	-0.19***	(-5.74)
MiddleIcome	0.13	0.34	0.10	0.31	0.03	(0.93)
HighIcome	0.81	0.39	0.65	0.48	0.16***	(4.00)
Northeast	0.87	0.34	0.87	0.34	0.00	(0.06)
Northwest	0.01	0.12	0.02	0.14	-0.01	(-0.49)
Center	0.04	0.20	0.08	0.28	-0.04	(-1.86)
South&Islands	0.06	0.23	0.07	0.26	-0.02	(-0.66)
NearbyHighSchool	0.88	0.32	0.86	0.35	0.02	(0.68)

Panel B.1.: Dichotomous Variable for treated

	(1) MALE		(2) FEMALE		(3) T-TEST $(\overline{x_m} - \overline{x_f})$	
	mean	sd	mean	sd	b	t
Pass	0.71	0.46	0.65	0.48	0.06	(1.48)
Retaker	0.26	0.44	0.33	0.47	-0.07	(-1.73)
cred40	0.56	0.50	0.44	0.50	0.12**	(2.83)
native	0.92	0.28	0.89	0.32	0.03	(1.12)
LowIcome	0.18	0.38	0.24	0.43	-0.06	(-1.72)
MiddleIcome	0.07	0.25	0.06	0.25	0.00	(0.04)
HighIcome	0.75	0.43	0.69	0.46	0.06	(1.56)
Northeast	0.84	0.37	0.87	0.34	-0.03	(-1.01)
Northwest	0.02	0.13	0.02	0.14	-0.00	(-0.35)
Center	0.06	0.25	0.06	0.24	0.01	(0.24)
South&Islands	0.10	0.29	0.07	0.26	0.02	(0.95)
NearbyHighSchool	0.84	0.36	0.88	0.32	-0.04	(-1.29)

Source: self-elaboration on primary & administrative data.
Notes: t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
[A] TOLC = has changed its composition for students enrolled from 2017 onwards (the English evaluation was introduced).
[B] variable not homogeneous in the sample (missing not at random) it is detectable only for students enrolled from 2017 onwards
[C] I-rat score could exist even if Treat is zero. It belongs to students who participated at TBL without reaching the minimum treatment dosage.
Table 2 – Descriptive statistics

This condition is verified for both male and female students and could suggest that students who self-select into treatment by accepting to attend the Team-Based learning are the most deeply motivated. Regardless of the cluster, the majority of students attended a secondary school in the same region as the university (neighbourhood proxy) and have a stable household financial situation. In the last column of Table 2 (panel A and Panel B), a t-test (Two-sample t-test with equal variances) on covariates was also included to see if the covariates assume significant gender differences within Treat groups.

T-test does not reveal any particular gender differences in the covariates except for TOLC scores (ComprehensionAbility, MathAbility, LogicAbility). Descriptive statistics show that, although female students score significantly worse on the entrance test, they manage to achieve an academic performance in the same line or even higher than that of male students. The worse performance in the entrance test by female students could also be due to the type of question framing. Previous literature has already emphasized that multiple-choice type questions result in disadvantages for females (Reardon et al. 2018) and that the higher risk perceived on average by female students in answering this type of question leads to their lower performance (Baldiga 2014; Karimi and Biria 2017).

Results of the Econometric models

i) Regression model

Table 3 illustrates the STATA outcome for models 1, 1a and 1b. Before commenting on them, it is important to remember that the regression is regarded as a dependent variable in which all fail marks (which are unobserved) were set at 17. The regression in Table 3 shows that participation in TBL is highly significant for all students and exam marks are 1.7 points higher than for those who do not participate. The main benefits are for female students (in line with the literature surveyed in Section 2), who achieved almost 3 more points in their exams by taking part in TBL. Being born in Italy seems to be important for male students' marks in macroeconomics, but not relevant for those of female students. The positive and significant impact of the Highlyquantitative, Slightlyquantitative and Nonquantitative variable coefficients are expected as the level of preparation and ability of the students is linked to their performance in all past exams; the impact, however, is not higher for courses with a high quantitative content. The positive and significant coefficient of Retaker, a variable that takes the value of one when the exam has been retaken, can be found only for female students; this result may be linked to a strategy that is more frequent with female students of sitting an exam as an attempt to acquire familiarity with the exam structure and then only accepting only the higher mark and this would require greater investigation (also interacting the variable with the TBL experience to test the positive impact of TBL on re-takers' achievements detected in the literature surveyed in Section 2).

Table 3 – Results of the estimation of models 1, 1a and 1b.

	(1) ALL	(2) FEMALE	(3) MALE
Treat	1.689***	2.809***	1.341*
	(3.52)	(3.77)	(2.11)
Female	-0.0286		
	(-0.08)		
Credits	0.0509***	0.0676***	0.0323
	(3.91)	(3.90)	(1.65)
Native	1.241	-0.701	3.388**
	(1.95)	(-0.94)	(3.17)
LowIncome	-0.173	0.426	-0.0604
	(-0.35)	(0.69)	(-0.08)
Highlyquantitative	0.251***	0.181	0.274**
	(3.32)	(1.65)	(2.63)
Slightlyquantitative	0.363***	0.536***	0.332**
	(4.80)	(4.79)	(3.13)
Nonquantitative	0.449***	0.428**	0.474**
	(4.18)	(2.77)	(3.17)
Retaker	0.814	1.677**	0.497
	(1.94)	(2.92)	(0.82)
NearbyHighSchool	-0.703	0.379	-1.512
	(-1.27)	(0.50)	(-1.86)
ComprehensionAbility	-0.0113	0.262*	-0.228*
	(-0.14)	(2.18)	(-1.99)
MathAbility	0.0525	0.0717	0.0398
	(0.81)	(0.77)	(0.44)
LogicAbility	-0.0160	0.0379	-0.0462
	(-0.21)	(0.38)	(-0.39)
Session FE	YES	YES	YES
Constant	-4.677*	-8.096*	-4.410
	(-2.11)	(-2.50)	(-1.42)
N	585	243	342

Source: self-elaboration on primary & administrative data. Notes: t statistics in parentheses
* p < 0.05, ** p < 0.01, *** p < 0.001

ii) Probit model

The outcome of probit estimations 2, 2a and 2b are displayed in Table 4 where marginal effects computed at the means of the variables are displayed. Once again, the diversity of results by gender can be observed. TBL treatment seems to have a positive and significant impact on male students' pass probability. In addition, it is interesting to highlight that having participated in the TBL seems to be the only variable determining the probability of males passing. This means that participating in the TBL becomes more important than the males' abilities (Highlyquantitative, Slightlyquantitative, Nonquantitative) and their diligence (number of credits acquired in the first year).

Table 4 – Marginal effects at means of model 2, 2a and 2b.

	(1) ALL	(2) FEMALE	(3) MALE
Treat	0.286	-0.260	0.642**
	(1.67)	(-0.82)	(2.83)
Female	0.144		
	(0.99)		
Credits	0.0177***	0.0250**	0.0142*
	(3.70)	(3.15)	(2.01)
Native	0.180	-0.339	0.705
	(0.81)	(-1.01)	(1.92)
LowIncome	-0.198	-0.112	-0.172
	(-1.10)	(-0.41)	(-0.59)
Highlyquantitative	0.0157	-0.001	0.0307
	(0.57)	(-0.02)	(0.83)
Slightlyquantitative	0.0638*	0.134*	0.0594
	(2.21)	(2.47)	(1.49)
Nonquantitative	0.123**	0.208**	0.0657
	(2.92)	(2.67)	(1.15)
Retaker	0.223	0.766**	0.006
	(1.45)	(2.83)	(0.03)
NearbyHighSchool	-0.216	-0.225	-0.473
	(-0.90)	(-0.58)	(-1.16)
ComprehensionAbility	0.00303	0.141*	-0.0659
	(0.09)	(2.29)	(-1.48)
MathAbility	0.0377	0.0189	0.0480
	(1.45)	(0.39)	(1.38)
LogicAbility	0.0532	0.0380	0.0825
	(1.79)	(0.78)	(1.85)
SESSION FE	YES	YES	YES
Constant	-4.802***	-8.074***	-3.668**
	(-5.08)	(-4.41)	(-2.87)
N	585	243	334

Source: self-elaboration on primary & administrative data.

Note: Statistical significance at the 1%, 5% and 10% levels is denoted by ***, **, *

Note: dy/dx for factor levels is the discrete change from the base level.

iii) Cragg's model (two-part, hurdle model)

Finally, Table 5 shows the output of Cragg's model presented in the equations 3, 3a and 3b.

As Cragg's model develops in two consecutive stages, the output of Table 5 must be analyzed in light of the findings of the probability of being promoted in the Probit model (Table 4). Table 5 displays the marginal effect computed at means of covariates rather than coefficients (β) because the latter indicates the estimate of the latent variable (Mark) whereas by giving an actual value (mean) to the covariate we can calculate the real Observing the first column of Table 5, we note that - once promoted - having participated in the TBL increases the examination marks by 3 points, a result validated for each level of significance. Especially the female subgroup seems to weigh on the sample size and effect. The largest effect in the table is seen for females who attended classes when TBL was implemented: among the female students who did not fail, those who participated in TBL scored approximately 6 points higher than those who did not. This evidence is significant for each level. On the contrary for male students, although positive, this variable is not significant. This result is in line with the literature results surveyed in Section 2 showing the higher impact of TBL on students' exam marks for certain groups of the students' population.

Table 5 – Output of models 3, 3a and 3b.

	(1) ALL	(2) FEMALE	(3) MALE
Treat	4.889**	7.624***	2.265
	(3.09)	(4.02)	(1.50)
Female	0.124		
	(0.14)		
Credits	0.101**	0.070*	0.059
	(2.66)	(2.07)	(1.42)
Native	2.384	-0.639	7.470*
	(1.46)	(-0.56)	(2.40)
LowIncome	0.359	1.650	0.040
	(0.29)	(1.52)	(0.02)
Highlyquantitative	0.355	0.404	0.276
	(1.72)	(1.86)	(1.15)
Slightlyquantitative	0.434*	0.468*	0.446
	(2.13)	(2.18)	(1.85)
Nonquantitative	0.952**	0.565*	1.016**
	(3.19)	(2.10)	(2.99)
Retaker	1.049	0.625	1.656
	(0.93)	(0.55)	(1.25)
NearbyHighSchool	0.349	2.045	-1.199
	(0.26)	(1.51)	(-0.75)
ComprehensionAbility	0.002	0.175	-0.269

	(0.01)	(0.86)	(-1.19)
MathAbility	-0.054	0.055	-0.088
	(-0.35)	(0.38)	(-0.50)
LogicAbility	-0.232	-0.026	-0.385
	(-1.20)	(-0.15)	(-1.58)
SESSION FE	YES	YES	YES
constant	-25.086*	-15.459*	-25.882*
	(-2.51)	(-2.12)	(-2.31)
sigma	4.577***	3.201***	4.146***
	(8.80)	(8.91)	(7.84)
N	346	143	203

Source: self-elaboration on primary & administrative data.
Notes: *t* statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Coefficients are omitted and marginal effects are displayed:
dydx(*) at means predict(e(17,.)), E(Mark| Mark>17), predict(e(17,.))

Conclusions and future research development

The primary aim of this essay is to evaluate the impact of Team-Based Learning on students' achievements measured by the exam marks and the probability of passing the exam with reference to different cohorts of students attending the Introductory Macroeconomics Course in a Bachelor's Degree Programme in a public University located in the North of Italy.

Multivariate analyzes performed in this essay provide evidence on H1 hypothesis that attending Team-Based Learning produces better learning outcomes measured by the exam grade both by estimating an OLS model and even more by estimating a Hurdle Model consistently with what has been shown in the literature: by analysing three different economics courses Espey (2022) found that TBL produces better learning outcomes together with higher levels of engagement in team activities and also Cagliesi and Ghanei (2022) found that TBL improved students' academic performance and reduced several achievement gaps in the economics class. The models' estimation provides evidence of a different impact by gender of the introduction of TBL thus satisfying the H2 hypothesis that students react differently to treatment depending on their gender. In fact, we find a higher impact, in terms of grades, for female students, with no significant impact on their probability of passing the exam, while, on the contrary, male students' probability of passing the exam is strongly affected by TBL, though bearing a lower effect on grades. The higher positive impact on female students' grades can be connected to the higher training that female students can benefit from while attending TBL due to the structure of TBL tests being more similar to the final exam structure. In fact, the latter contains multiple choice questions characterised on average by a lower performance for female students, as shown by Griselda (2020); Karimi and Biria (2017), Baldiga (2014). Meanwhile, the two joint results (impact on grades and probability of passing) might suggest that the practice of TBL not only influences knowledge but can also affect students' behaviour. The literature shows that female students, who are more risk-averse, tend to show up for the exam only if they are fully prepared, while male students, in contrast, tend more to "try it" (Dohmen, 2010; Croston, 2009; Niederle, 2007). Therefore, we can claim that the TBL practice on both male and female students may have also acted on their approach to exams by empowering – or at least compensating – them.

With regards to the common belief that female students perform worse than male students in macroeconomics as a quantitative subject (H3 = Female performance in Macroeconomics is lower), the hypothesis is not confirmed by our results. In fact, although descriptive statistics (Table 2, panel A) show that women score significantly lower on the entrance test (TOLC and any of its subgroups), they manage to achieve similar performance in line with their male counterparts (or even higher if they attended TBL). Once again, one possible explanation for this discrepancy is that the TOLC test is based on multiple-choice questions, in which female students on average tend to underperform. Supporting this hypothesis is the fact that female students also underperform in verbal comprehension in the entry test (TOLC), whereas they usually tend to be more talented in this latter. These results are remarkable as they suggest that TBL practice may help female students in overcoming their disadvantage in multiple-choice questions even in quantitative domains. This point is another research goal that we are trying to further qualify by analysing the structural break of the Covid pandemic that caused the massive use of this testing modality. With regards to the H4 hypothesis that TBL could help in overcoming gender differences in macroeconomics, even if our analysis does not show significant gender differences in macroeconomics exam marks, participating in TBL, as tested by our estimation, has an extremely positive impact on female outcomes and this can help to prevent the gap from widening. In addition, as proved by the models estimated and referred to above, TBL teaching may have affected not only the economics grade but also female students' behaviours in approaching multiple-choice tests. Observed changes in students' behaviour after TBL practices have also been detected by other authors, including Christensen et al. (2019); Lin (2019); Dearnley et al. (2018). Collectively, these findings indicate a positive and meaningful correlation between TBL course attendance and performance in macroeconomics exams, while accounting for individual variables related to students' socio-demographic and cognitive skills. Additionally, these results reveal a shift in students' behaviour and exam strategies, which warrants further investigation through data collection on this dimension.

These findings have important implications for policy interventions aimed at improving macroeconomic education outcomes or – in general – mitigating gender imbalances in quantitative courses and/or multiple-choice evaluation. Given the positive association between attending TBL courses and exam performance, institutions are encouraged to promote the use of TBL in Macroeconomics courses. This may involve organising training for lecturers on how to implement this teaching methodology effectively. Furthermore, universities could offer financial incentives or other forms of support to encourage teaching staff to adopt TBL, thereby promoting gender equality and improving learning outcomes for all students.

Moreover, they could provide training and resources to support female students in developing the skills and confidence needed to perform better on multiple-choice tests. This last policy could also be pursued in secondary schools in order to impact on female students' performance in multiple-choice tests that usually characterise entrance tests for university courses with a limited number of admitted students. However, we are aware of the limitations of the present study. The TBL course attendance is not compulsory, and students can opt out and not attend the TBL sessions or not attend the course but simply sit the exam as not-attending students. However, the participation rate is very high in our sample. Most of the students who had the option of attending the TBL course did decide to opt in, although we could not exclude students' self-selection into treatment even if it is conceivable to assume that the covariates included in the analyzes help to correct for bias due to self-selection, but a stronger counterfactual group is needed. Further developments include the introduction of a parallel traditional course in Introductory Macroeconomics held by another Lecturer by following a lecture-based approach without any opportunity to have TBL sessions, to improve the evaluation of the impact of TBL through the counterfactual or diff-in-diff techniques. Alternatively and/or additionally, we suggest including the Heckman (1979) correction for non-random selection in the treatment. Finally, the application of the Oaxaca decomposition (Blinder 1973; Oaxaca 1973) proposed by Bauer and Sinning (2010) may allow the detection of gender differential in Introductory Macroeconomics, not due to differences in the observed characteristics. This method is preferred

over the one implemented in Jann (2008) because Monte Carlo simulations demonstrate that in the case of censored dependent variables, this decomposition method produces more reliable results than the conventional Blinder–Oaxaca decomposition for linear regression models (Bauer and Sinning, 2010).
Further developments recognise the limits related to the simple use of test score and the necessity to include the evaluation of others more subjective indicators as students' interest, enjoyment, self-efficacy, etc… (Addabbo, Di Tommaso, Maccagnan 2014). Moreover, viewing TBL also as an inclusive teaching methodology, we are planning to measure students' perceived sense of inclusion by including in our data collection validate scales regarding the "sense of belonging" (Good et all. 2012) and Survey on Diversity, Equity and Inclusion. Finally, further research could explore the potential effectiveness of TBL in other subject areas and courses, comparing it with other teaching techniques and their respective impact on students' learning outcomes.

References

Abio, Gemma, Alcaniz Manuela, Gomez-Puig Marta, Rubert Gloria, Serrano Monica, Stoyanova Alexandrina, and Vilalta-Bufi, Montserrat. 2019. "Retaking a course in economics: Innovative teaching strategies to improve academic performance in groups of low-performing students." Innovations in Education and Teaching International 56 (2): 206–16. https://doi.org/10.1080/14703297.2017.1389289

Addabbo, Tindara, Di Tommaso Maria Laura and Maccagnan Anna. 2014. "Gender Differences in Italian Children's Capabilities." Feminist Economics 20 (2): 90–121. https://doi.org/10.1080/13545701.2013.844846

Baldiga, Katherine. 2014. "Gender Differences in Willingness to Guess." Management Science 60 (2), 434-448. https://doi:10.1287/mnsc.2013.1776

Ballard, Charles, and Marianne Johnson. 2005. “Gender, Expectations, and Grades in Introductory Microeconomics at a US University.” Feminist Economics https://doi.org/10.1080/1354570042000332560

Ballester Pilar, Carmen. “Factors That Influence Academic Performance: Analyzing Gender Differences in Accounting Students”. Revista Educación 43, no. 2 (2019): 29. https://doi.org/10.15517/revedu.v43i2.28916

Bauer, Thomas, and Mathias Sinning. 2010. «Blinder-Oaxaca decomposition for Tobit models». Applied Economics 42, n. 12: 1569–75. https://doi.org/10.1080/00036840701721612

Bellini, Claudia, De Santis Annamaria, Sannicandro Katia, and Cecconi Luciano. 2020. “Dalla formazione dei docenti alle competenze trasversali degli studenti: un progetto di faculty development presso Unimore” in Faculty Development in Italia. Valorizzazione delle competenze didattiche dei docenti universitari, a cura di A. Lotti, P. A. Lampugnani.

Blinder, Alan. 1973. “Wage Discrimination: Reduced Form and Structural Estimates”. Journal of Human Resources 8 (4): 436–55.

Bilen, Eren and Matros, Alexander. 2021. “Online cheating amid COVID-19.” Journal of Economic Behavior & Organization 182, 196-211. https://doi.org/10.1016/j.jebo.2020.12.004

Borrowman, Mary, and Stephan Klasen. 2020. “Drivers of Gendered Sectoral and Occupational Segregation in Developing Countries.” Feminist Economics 26 (2): 62–94. https://doi.org/doi:10.1080/13545701.2019.1649708

Burke, William. 2009. “Fitting and Interpreting Cragg's Tobit Alternative Using Stata.” The Stata Journal: Promoting Communications on Statistics and Stata 9 (4): 584–92. https://doi.org/10.1177/1536867X0900900405

Cagliesi, Gabriella and Ghanei Mahkameh. 2022 “Team-Based Learning in Economics: Promoting Group Collaboration, Diversity and Inclusion”, Journal of Economic Education, 53(1), pp. 11–30.

Clerici-Arias, Marcelo. 2020. “Underrepresented groups in the economics major. The impact of using team-based learning in the principles course.” Paper presented at the 2020 Annual Meeting of the Allied Social Science Associations, San Diego, CA, January 4.

Coyne, Leanne, Takemoto Jody K., Parmentier Brittany, Merritt Thayer, and Sharpton Rachel. 2018. “Exploring Virtual Reality as a Platform for Distance Team-Based Learning.” Currents in Pharmacy Teaching and Learning 10 (10): 1384–90. https://doi.org/10.1016/j.cptl.2018.07.005

Christensen, Jacqueline, Harrison Jennifer, Hollindale, Janice, and Wood, Kayleen. 2019. “Implementing team-based learning (TBL) in accounting courses.” Accounting Education, 28(2), 195-219.

Croson, Rachel, and Gneezy, Uri. 2009. “Gender differences in preferences.” Journal of Economic literature, 47(2), 448-474.

Dearnley, Chris, Rhodes Christine, Roberts Peter, Williams Pam, and Prenton, Sarah. 2018. “Team based learning in nursing and midwifery higher education; a systematic review of the evidence for change.” Nurse Education Today, 60, 75-83.

De Santis, Annamaria, Sannicandro Katia, Bellini Claudia, and Cecconi Luciano. 2019. “Valutazione e soft skills nella didattica universitaria, Training actions and evaluation processes.” Atti del Convegno Internazionale SIRD, pp. 493-502.

Di Tommaso, Maria Laura, Maccagnan Anna, and Mendolia Silvia. 2021. “Going Beyond Test Scores: The Gender Gap in Italian Children’s Mathematical Capability.” Feminist Economics 27:161–187. https://doi.org/10.1080/13545701.2021.1908574

Dohmen, Thomas, Falk Armin, Huffman David, and Sunde Uwe. 2010. “Are Risk Aversion and Impatience Related to Cognitive Ability?”

American Economic Review 100 (3): 1238–60. https://doi.org/10.1257/aer.100.3.1238

Espey, Molly. 2022 "Variation in Individual Engagement in Team-Based Learning and Final Exam Performance", International Review of Economics Education, 41.

Espey, Molly. 2018. "Diversity, effort, and cooperation in team-based learning." Journal of Economic Education. 49 (1): 8–21.

Espey, Molly. 2012. "Enhancing Critical Thinking in Economics Using Team-Based Learning." Conference Paper/ Presentation in Agricultural and Applied Economics Association (AAEA). https://doi.org/10.22004/AG.ECON.123521

Griselda, Silvia. 2020. "Different Questions, Different Gender Gap: Can the Format of Questions Explain the Gender Gap in Mathematics?" Job Market Papers pg 710.

Goh, Cheng Siew, Heap-Yih Chong, Lynne Jack, and Adam Fuad Mohd Faris. 2020. "Revisiting Triple Bottom Line within the Context of Sustainable Construction: A Systematic Review." Journal of Cleaner Production 252: 119884. https://doi.org/10.1016/j.jclepro.2019.119884

Good, Catherine, Rattan Aneeta, Dweck, Carol. 2012. "Why do women opt out? Sense of belonging and women's representation in mathematics." Journal of Personality and Social Psychology, 102(4), 700–717. https://doi.org/10.1037/a0026659

Heckman, James. 1979. "Sample Selection Bias as a Specification Error", Econometrica, Vol. 47, No. 1, 153–161.

Hettler, Paul. 2015. "Student demographics and the impact of team-based learning." International Advances in Economic Research 21 (4): 413–22. https://doi.org/10.1007/s11294-015-9539-7

Imazeki, Jennifer. 2015. Getting students to do economics: An introduction to team-based learning. International Advances in Economic Research 21 (4): 399–412.

Jakobsen, Krisztina, and Knetemann, Megan. 2017. “Putting Structure to Flipped Classrooms Using Team-Based Learning.” International Journal of Teaching and Learning in Higher Education 29 (1), 177-185. ISSN: EISSN-1812-9129.

Jann, Ben. “The Blinder–Oaxaca Decomposition for Linear Regression Models.” The Stata Journal: Promoting Communications on Statistics and Stata 8, n. 4 (2008): 453–79. https://doi.org/10.1177/1536867X0800800401

Karimi, Mehrnoush, and Biria, Reza 2017. “Impact of Risk Taking Strategies on Male and Female EFL Learners’ Test Performance: The Case of Multiple Choice Questions.” Theory and Practice in Language Studies 7 (10), 892-899. https://doi:10.17507/tpls.0710.10

Lin, Jian-Wei 2019. “The impact of team-based learning on students with different self-regulated learning abilities.” Journal of Computer Assisted Learning, 35(6), 758-768. https://doi.org/10.1111/jcal.12382

Michaelsen, Larry K., Davidson Neil, and Major Claire. 2014. “Team-Based Learning Practices and Principles in Comparison with Cooperative Learning and Problem-Based Learning.” Journal on Excellence in College Teaching. https://www.semanticscholar.org/paper/Team-Based-Learning-Practices-and-Principles-in-and-Michaelsen-Davidson/cf017faa464c6a7f0ce4707e14242fcbd667959f

Michaelsen, Larry K., and Michael Sweet. 2008. “The Essential Elements of Team-Based Learning.” New Directions for Teaching and Learning (116): 7–27. https://doi.org/10.1002/tl.330

Michaelsen, Larry K., Fink Arletta, and Bauman, L. Dee. (2004a). Centers for Teaching and Technology. Available at https://bit.ly/3sJzYSQ.

Michaelsen, Larry K., Fink Arletta, and Bauman, L. Dee. (2004b). "Team-based learning: A transformative use of small groups in higher education." Sterling, VA: Stylus.

Miller, Jane E., Yana Van Der, and Meulen Rodgers. 2008. "Economic Importance and Statistical Significance: Guidelines for Communicating Empirical Research." Feminist Economics 14 (2). Routledge: 117–149. https://doi.org/10.1080/13545700701881096

Niederle, Muriel, and Vesterlund, Lise. 2007. "Do women shy away from competition? Do men compete too much?" The quarterly journal of economics, 122(3), 1067-1101.

Nguyen, Joseph G., Keuseman, Kristopher J., and Humston, Jonathan J. 2020. "Minimize Online Cheating for Online Assessments During COVID-19 Pandemic." Journal of Chemical Education 97 (9), 3429-3435. https://doi:10.1021/acs.jchemed.0c00790

Odell, Kathleen E. 2018. "Team-Based Learning and Student Performance: Preliminary Evidence from a Principles of Macroeconomics Classroom." International Review of Economics Education: 44–58. https://doi.org/10.1016/j.iree.2018.01.001

Oaxaca, Ronald. 1973. "Male-Female Wage Differentials in Urban Labor Markets." International Economic Review 14, no. 3: 693. https://doi.org/10.2307/255981

Parappilly, Maria, Woodman, Richard John, and Randhawa, Sharmil. 2021. "Feasibility and Effectiveness of Different Models of Team-Based Learning Approaches in STEMM-Based Disciplines." Res Sci Educ 51 (Suppl 1), 391–405.

Parmelee, Dean X., and, Michaelsen Larry K., 2010. "Twelve tips for doing effective Team-Based Learning (TBL)." Medical Teacher 32 (2), 118-122. https://doi:10.3109/01421590903548562

Reardon, Sean F., Kalogrides Demetra, Fahle, Erin M., Podolsky Anne, and Zárate, Rosalia. 2018. "The Relationship Between Test Item Format and Gender Achievement Gaps on Math and ELA Tests in Fourth and

Eighth Grades." Educational Researcher 47 (5). https://doi:10.3102/0013189X18762105

Ruder, Phil., Maier Mark H., and Simkins, Scott Paul. 2021. "Getting started with team-based learning (TBL): An introduction." Journal of Economic Education, 52(3), pp. 220–230.

Simkins, Scott Paul, Maier Mark H., and Ruder, Paul. 2021 "Team-based learning (TBL): Putting learning sciences research to work in the economics classroom." Journal of Economic Education, 52(3), pp. 231-240.

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

Tobin, James. 1958. "Estimation of Relationships for Limited Dependent Variables". Econometrica 26, no. 1: 24–36. https://doi.org/10.2307/1907382

Watson, Warren E., Michaelsen, Larry K., and Sharp, Walt. 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

Acknowledgments
We thank the Introductory Macroeconomic Course instructor and tutors, the teaching coordinator of the Department and the responsible of the administrative data sets for the assistance provided in administrative data retrieval. We thank an anonymous referee for the stimulating comments received on a previous version of the paper as well as the participants to the International Conference "Innovative and Inclusive Academia" held at the Department of Economics Marco Biagi, University of Modena & Reggio Emilia, Italy on October 19th and 20th 2023 and to the 5th National Conference on Faculty Development "From policies to practices. Teaching professionalism in institutional and technological evolution" held at the University of Genoa on June 13th and 14th 2024 for their insightful comments. Funding from the Gender Equality in Academia FAR 2024 University of Modena and Reggio Emilia Department of Economics Marco Biagi Research Fund project - is gratefully. The

responsibility for the content of the paper and any errors remains exclusively with the authors.

APPENDIX

Table A1 – Description of the main variables

Variable	Name of the variable	Definition
Dependent variable		
OUTCOME IN MACROECONOMICS	Mark	Continuous variable which reports the students' verbalized grade in Macroeconomics. It ranges from 18 to 30.
OUTCOME IN MACROECONOMICS	Pass	Dummy variable equal to 1 if the student passes the exam and to 0 if he/she fails.
Independent variables		
EFFECTIVE PARTICIPATION AT TBL	Treat	Dummy variable equal to 1 if the student participated in at least 5 over 6 Team-Based Learning lessons and 0 otherwise. [participation rate higher than 80%]
FEMALE	Female	Dummy variable equal to 1 if the student is a female and equal to 0 if they are a male.
PERIOD CONTROL	Session	A set of 17 dummy variables which take value 1 in correspondence with one of each 17 different periods 0 for the remaining.
COMPLIANCE AT THE END OF THE FIRST YEAR	Credits	Continuous variable equal to the credits that the student earned in the first year. It ranges from 0 to 60 and we considered it important because the macroeconomics course is held in the following year (in the second year).
NATIVE	Native	Dummy variable equal to 1 if the student was born in Italy and 0 otherwise.
PREVIOUSLY FACE THE EXAM	Retaker	Dummy variable which has a value of 1 if the student is repeating the exam and 0 if the student is attempting the exam for the first time.
UNIVERSITY PERFORMANCE [see Table A2 for more details]	Highlyquantitative Slightlyquantitative Nonquantitative	Continuous variable which computes the mean of students on exams which have a high quantitative content. Continuous variable which computes the mean of students on exams which have a medium quantitative content. Continuous variable which computes the mean of students on exams which do not have a quantitative content.
NEIGHBOURHOOD	NearbyHighSchool	Dummy variable which has a value of 1 if the student attended a secondary school in the same region as the university and 0 otherwise.
UNIVERSITY ENTRANCE SCORE IN READING COMPREHENSION	ComprehensionAbility	Continuous variable which reports students' performance in reading comprehension at TOLC. The result is determined by the number of correct (1 point), wrong (-0.25 point) and not given answers (0 points).

UNIVERSITY ENTRANCE SCORE IN MATHS	MathAbility	Continuous variable which reports students' performance in maths at TOLC. The result is determined by the number of correct (1 point), wrong (-0.25 point) and not given answers (0 point).
UNIVERSITY ENTRANCE SCORE IN LOGIC	LogicAbility	Continuous variable which reports students' performance in logic at TOLC. The result is determined by the number of correct (1 point), wrong (-0.25 point) and not given answers (0 point).
Minor variables – Used for descriptive statistics or to give an in-depth view of the sample		
INTERVENTION DOSAGE	Dosage	Continuous variable which ranges from 0 to 6 and considered students' participation at TBL lessons. Dosage was computed by counting (and summing) each Irat score when it was not missing.
INDIVIDUAL PERFORMANCE AT I-RAT	IratScore	Continuous variable which computes the mean of all Irat scores collected by students.
EXAM ATTEMPTS NUMBER	Attempts	Continuous variable which indicates the number of times the student takes the exam (1 for the first Attempts and progressive number for further tries).
ECONOMIC HARDSHIP INDICATOR	LowIncome; MiddleIncome; HighIncome	Three dummy variables equal to 1 if student's family unit has an equivalent economic status indicator respectively: lower than €23,000 for LowIncome ; lower than €45,000 Middle income or higher than €45,000 for HighIncome.
ENROLMENT WAITING PERIOD	EnrollGap	Continuous variable that corresponds to years from completing secondary school to university enrolment.
NEIGHBOURHOOD (NUTS1 ARRANGEMENTS)	Northeast Northwest Centre South&Islands	Four Dummy variable which takes the value of 1 if the student attended a high school in one of those macro areas (NUTS1) and 0 for the others.

Table A2 – Disaggregation of macro-structures of examination performance

Pane A - Highly Quantitative domain
Economics of financial intermediaries
Monetary economics
Corporate finance, financial analysis
Corporate finance
Introduction to microeconomics
Macroeconomics
Mathematics for economics and finance
Financial and actuarial mathematics
Microeconomics
Models for financial investments
Mathematics and financial mathematics
Savings and financial choices of enterprises
Financial science
Statistics
Panel B - Slightly Quantitative domain
Business economics
Business economics 2
Economics of credit companies
Economics of credit companies
Securities market economics
Welfare systems
Panel C - Non-quantitative domain
Marketing
Industrial law
Labour law
Public law
Private and commercial law
Tax law
European Union law
Economics and business management
International economics
Economics and institutions of industrial districts
Economics and labour policies
Ethics and corporate social responsibility
EU integration and community policies
International marketing I
International marketing II
Business organisation
Programming and control
Marketing research
Economic history
Italian economic history
Economic history

Table A3 – Correlation with MARKS and key relationships

	CONTROLS (no TBL)		TREATED (TBL)	
	(1) Marks for Male	**(2) Marks for Female**	**(3) Marks for Male**	**(4) Marks for Female**
Dosage	-0.03[1]	-0.13[1]	-0.01	0.14*
IratScore	-0.04[1]	-0.14*[1]	0.24***	0.30***
Highlyquantitative	0.39***	0.53***	0.48***	0.45***
Slightlyquantitative	0.38***	0.55***	0.41***	0.60***
Nonquantitative	0.48***	0.61***	0.46***	0.43***
TOLC	0.15*	0.33***	0.27***	0.33***
ComprehensionAbility	0.04	0.38***	0.17**	0.09
MathAbility	0.09	0.19*	0.25***	0.40***
LogicAbility	0.10	0.16	0.23***	0.16*
EnglishAbility	0.16	0.04	0.14*	0.15*
Attempts	-0.16**	-0.18**	-0.16**	-0.16*
GapfromDiploma	0.08	0.00	-0.11	0.01
Credits	0.41***	0.58***	0.35***	0.48***

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

Notes: The entire output of the correlation matrix (all the pairwise correlation coefficients between the 14 variables = a 14×14 matrix for each group) is omitted and only the column concerning the correlations with the dependent variable is reported **(Mark**)).

[1] These variables are also present for the control group because not all tbl participants received the minimum dosage to be considered treated.

Addendum A4 - Fit test: Simple Tobit VS Craggs Model

Eq 4 ALL $$\lambda = 2 \cdot [-262{,}10 + (-814{,}72) - (-1294{,}55)] = 435{,}43$$

Eq 4 FEM $$\lambda = 2 \cdot [-94{,}49 + (-306{,}8) - (-499{,}51)] = 196{,}44$$

Eq 4 MALE $$\lambda = 2 \cdot [-145{,}89 + (-472{,}01) - (-765{,}61)] = 245{,}41$$

The three values resulting from the above formulae all exceed the chi-square threshold for 30 or 29 degrees of freedom (covariates plus the intercept) of the equations.