UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Dipartimento di Economia Marco Biagi

DEMB Working Paper Series

N. 219

Team-Based Learning as an innovative teaching methodology: assessing gender inclusiveness in a quantitative economic subject

Tindara Addabbo[1], Chiara Tasselli[2], Filippo Damiani[3]

December 2022

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

ISSN: 2281-440X online

Team-Based Learning as an innovative teaching methodology: assessing gender inclusiveness in a quantitative economic subject

Tindara Addabbo*
Chiara Tasselli**
Filippo Damiani***

Department of Economics Marco Biagi Working Papers Series – December 2022 No. 219

*Department of Economics Marco Biagi, University of Modena and Reggio Emilia & Gender Equality Observatory - Marco Biagi Foundation

**Phd Candidate Labour, Development and Innovation, Department of Economics Marco Biagi and Marco Biagi Foundation

***Postdoctoral Research Fellow, Department of Economics Marco Biagi

Keywords: inclusive and innovative teaching, Team-Based Learning; Gender equality; higher education; SDG5

JEL codes: I23, I24, J16

TABLE OF CONTENT

ABSTRACT

This paper analyses the impact on students' performance of the introduction of team-based learning (TBL, for short) teaching methodology. TBL has proven to be a powerful and versatile teaching strategy that enables teachers to take small group learnings to a new level of effectiveness according to the empirical evidence of its implementation worldwide since 2020.

The analysis involves various cohorts of macroeconomics students at the University of Modena and Reggio Emilia Marco Biagi Department of Economics (DEMB). We exploit the structural break in the academic year 2017/2018 in which TBL was introduced in the Macroeconomics Course at the second year of the Bachelor Programme in Business and management to test the impact of TBL teaching methods in learning outcomes also considering students' heterogeneity.

Within this perspective of inclusiveness, special attention is given to gender bias in TBL effectiveness and outcomes in Macroeconomics measured by using Macroeconomics exam grades. Comparisons between groups were made using econometric and statistical analysis on a rich and self-constructed database. The latter, composed using multiple administrative and primary data sources, allows controlling for student socio-demographic characteristics and their academic careers. The econometric analysis starts with a *multivariate regression* to estimate the TBL's effect on grades in macroeconomics, then moves to a *probit analysis* that returns the probability of passing the exam and concludes with a *Cragg model* (two-part Hurdle model) which offers improved estimates by correcting for the censored dependent variable. Regardless of gender, a positive impact of TBL on students' performance is detected. Treated, especially if female, improves macroeconomic scores, and this is consistent with the literature reviewed. On the other hand, males benefit more from the treatment through a significant increase in the likelihood of passing the exam.

1. INTRODUCTION

Previous studies in the education field have shown that gender plays a decisive role in students' academic enrolment and performance in economics, scientific and technological universities. Even if empirical evidence (Castagnetti & Rosti 2009) and data[1] show that women outperform men in Italian Universities, female low self-confidence in quantitative and scientific courses is found to affect their access to employment and horizontal segregation in the labour market. That structural problem -in addition, to perpetuating gender inequality in this field and making 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 has been 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 through its application to economics has been limited (Cagliesi & 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 the students' soft skills in problem-solving and teamwork development.

As assessed by Simkins, Maier, & Ruder (2021), TBL intentionally promotes learning strategies that learning sciences research identified as highly effective to create powerful learning environments for students. The attention paid to the group's composition resulting in within groups diversity also on the ground of gender allowed us to test also its impact on inclusion.

The analysis compares 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[2] attending a macroeconomics course at the Marco Biagi Department of Economics of the University of Modena & Reggio Emilia.

It can be, therefore hypothesized that:
H_1 = *attending Team-Based Learning Lessons produces better learning outcomes.*
H_2 = *students react differently to treatment depending on their gender.*
H_3 = *Female performance in Macroeconomics is lower.*
H_4 = *TBL could help overcome gender differences in macroeconomics.*

We do try to solve 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).

The paper is structured as follows: Section 2 is dedicated to the literature review and presents the principles of Team-Based Learning that we adopted it in our courses. Section 3 presents the sample selection and distribution (3.1) together whit a detailed and in-depth insight into data and variables description (3.2) which continues in the annexes section. Then, the methodology followed and the results of the estimated models are presented in section 4. Section 5 provides some concluding remarks.

---

https://www2.almalaurea.it/cgi-php/universita/statistiche/tendine.php?LANG=it&config=profilo.
[2] For a total amount of 1,024 exam attempts.

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, specific soft-skills and, in its implementation, allows a high degree of inclusiveness: Team based learning (TBL).

TBL has been 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 & 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 the students to apply course contents since, during TBL activities, the development of self-managed learning teams is promoted (Michaelsen & Sweet, 2008; Michaelsen, Davidson & Major, 2014).

TBL teams composition and in 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, & 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 possible (Parmelee & Michaelsen, 2010; Phillips et al., 2008), which means that groups are deliberately formed to be diverse and cohesive (Kathleen & Odell 2018). The dimension of groups is of 5-7 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 as 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 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 teammates (Step 4); this last part is fundamental to enhance the ability to work together and positively contribute to the team (Michaelsen, Davidson & Major, 2014) and to avoid freeriding (Hettler, 2015).

While frequently implemented in a face-to-face classroom, TBL has received limited attention in the online learning environment were 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).

Amongst the positive impact of TBL, the literature has shown increased students' engagement both in class and out of class (Imazeki 2015; Espey, 2012; Ruder, Maier & Simkins 2021) and increased attendance (Abio et al. 2019). Evidence has been provided on a positive impact of the adoption of TBL in the percentage of show up 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 TBL implementation in principles of microeconomics and quantitative methods courses 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 score in TBL sections thus supporting the hypothesis that TBL can have a higher impact on groups that are typically disadvantaged. Cagliesi & 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 TBL methodology to foster inclusion, not only evidence has been provided of the reduction in the achievement gaps for minorities attending courses using TBL sessions, but also evidence has been provided on the TBL approach to be more attractive for female and non-white students (Clerici-Arias, 2021).

Another line of investigation on TBL evaluation concerns the impact of teams' characteristics on teams or individual outcomes or behaviour in teams on individual outcomes. Espey (2018) analyses what measurable characteristics of teams influence team and individual performance on 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 engagement in team-based activities.

3. TEAM BASED LEARNING IMPLEMENTATION AND THE GENERATED DATA

TBL methodologies have already been adopted in the University of Modena and Reggio Emilia (Unimore) in 2017 within the project "Didactics 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) Unimore identified the two soft skills that at the beginning of the project were the most demanded in the labour market: problem-solving, i.e. an approach to work that, by identifying priorities and critical issues, allows the identification of the best possible solutions to problems; teamwork, i.e. the willingness to work and collaborate with others, having the desire to build positive relationships aimed at achieving the assigned task. TBL has then been chosen as a methodology able to develop these soft skills and implemented in the academic year 2017/2018 in 16 courses with 16 control courses that have allowed an evaluation of the impact of TBL on students' soft skills. Instructors and tutors involved in the TBL courses have been involved in a training course to acquire knowledge on TBL methodology and on how to restructure their syllabus. A community of practices has then been built within Unimore in strict collaboration with the Italian National TBL Community of practices.

The undergraduate course in Macroeconomics analysed in this paper, has been involved from the very beginning of the TBL implementation and the data collected refer in total to 891 students (1,345 including those who repeated the exam) who attended the course from the academic year 2016/2017 (when TBL has 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 Prof. Malcolm K. Sparrow in 2011 to maximize the mixing of students across the class (Sparrow, 2011). The variables that have been used in this implementation of the GRumbler to form the TBL groups have been collected throughout a survey run before TBL classes in each academic year and refer to gender, age, origin, type of high school attended, grades in Math 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 has been kept permanent with semester-long teams.

The implementation of TBL in the Macroeconomics semester course is structured in 30 lectures using active learning techniques and six TBL units with partial pre-class assignments following the Readiness Assurance Process four steps structure described in Section 2.

3.1 Sample.

To avoid any possible contamination of the data with 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.[3] The final sample, therefore, consists of 711 students and 1,024 exams attempts.
Students are cohorts attending the second year of the Undergraduate Course in Macroeconomics from the academic year 2016/2017 till 2020/2021[4]. The lectures take place in the first Semester of the academic year from September to December. A total of 6 exams run each academic year: 2 in the Winter Session, 3 in the Summer Session and 1 in the Fall Session.
Table 1 shows the composition of the final sample also including the exam session undertaken. The table shows that before the introduction of TBL, students preferred to sit for the second winter call, whereas from 2018 onwards they show up predominantly to the first call of the Winter Session.
Out of the 1,024 examination tests, 439 were carried 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[5] in TBL nearly 5 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] Moreover, for the last exam sessions, it would not be possible to use the whole list of covariates proposed in the analysis as the databases are currently being updated.
[4] Only until the February session because the Summer sessions suffer from pandemic interference.
[5] Percentage gap estimated only on students who attended classes from the introduction of TBL and, hence, got the opportunity to participate.

Tab 1 – Sample by gender and date of exam

	MALE		FEMALE		
SESSION	*Treat=0*	*Treat=1*	*Treat=0*	*Treat=1*	TOTAL
Jan 17	33 100%	X	36 100%	X	69
Feb 17	76 100%	X	46 100%	X	122
July 17	16 100%	X	11 100%	X	27
Sept 17	11 100%	X	12 100%	X	23
Gen 18	19 28.8 %	47 71.2%	9 17.3%	43 82.7%	118
Feb 18	21 36.8%	36 63.2%	9 18.4%	40 81.6%	106
May 18	3 75%	1 25%	6 50%	6 50%	16
June 18	7 63.6%	4 36.4%	4 40%	6 60%	21
July 18	8 57.1%	6 42.9%	6 40%	9 60%	29
Jan 19	16 21.1%	60 78.9%	7 18%	32 82.1%	115
Feb 19	21 38.2%	34 61.8%	9 30%	21 70%	85
May 19	2 15.4%	11 84.6%	2 25%	6 75%	21
June 19	6 60%	4 40%	5 55.6%	4 44.4%	19
July 19	10 62.5%	6 37.5%	3 60%	2 40%	21
Sept 19	9 56.3%	7 43.8%	5 50%	5 50%	26
Jan 20	17 23%	57 77.0%	21 35.6%	38 64.4%	133
Feb 20	16 43.2%	21 56.8%	13 36.1%	23 64%	73
Total	291	294	204	235	1024

Source: self-elaboration on primary & administrative data.

3.2 Data and variables description:

To investigate the gender differences in macroeconomics among undergraduate students analysed, multiple data sources have been merged.
Administrative data have been downloaded for the purpose from the Unimore student management system[6], results of intermediate tests collected by the professor and socio-demographic and behavioural covariates obtained by submitting students a questionnaire.

[6] *Student Management System* (Sistema per la gestione studenti: ESSE3) is one of the "core" services of Cineca's suite of products to support "Didactics and Students" in the university environment. First of all ESSE3 allows to manage (and follow) the entire academic' *"life cycle"* of the student.

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):

The dependent variables used to represent students' academic performance are two: 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 have 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 30).

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[7] in Math *(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 a and iii) are considered proxies of ability before entering university.

iv) if they have already Attempted the test *(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) outside prescribed time students *(OverTimeGrad)*

viii) the average of all the exams taken by the students during their academic career subdivided into 3 macro-groups[8](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:

ix) gender *(Female)*

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

xi) *(Native)* Italian nationality

In addition, time-fixed effects account for all unobservable factors that are changing across sessions. This method is useful for increasing the adjusted R-squared because it allows each session to have a customised coefficient that increases the goodness of fit.

Table A1 in the Appendix provides an in-depth and detailed description of the variables grouped by:

- Dependent variables
- Independent variables
- Minor variables used for descriptive statistics or to provide an in-depth view of the sample

[7] TEST ONLINE CISIA: more detail in tab. 1 in addendum.

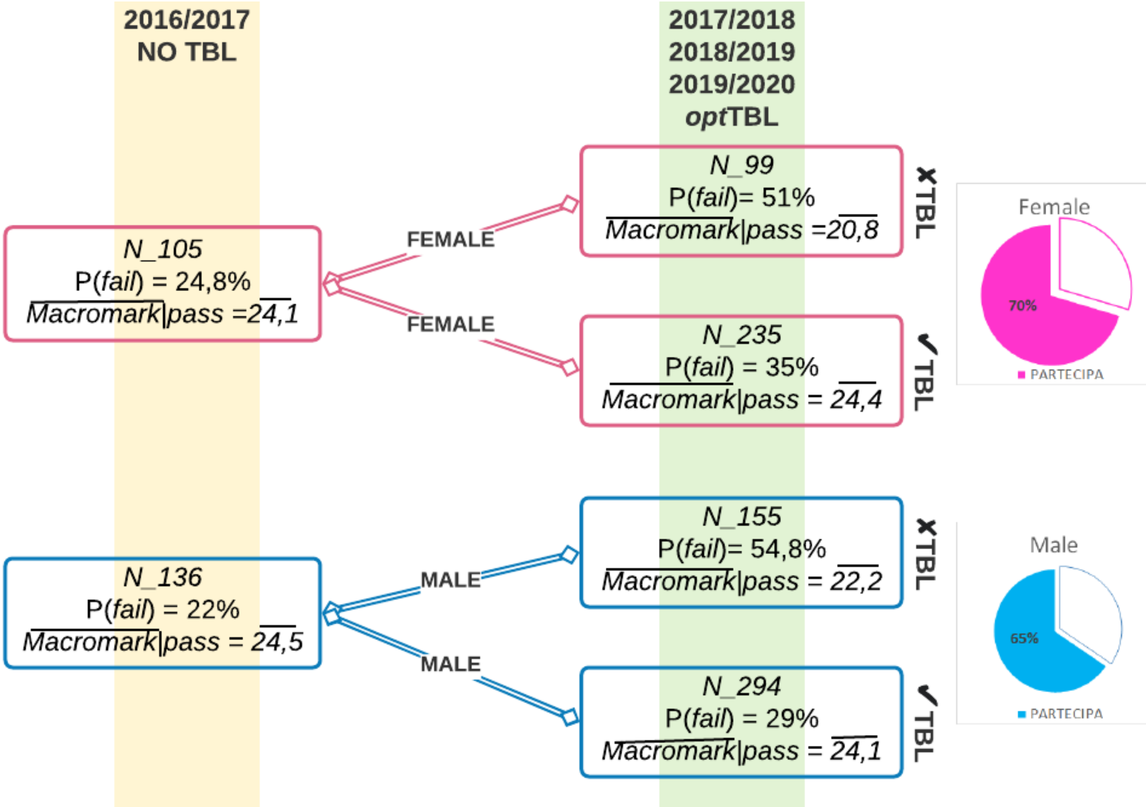
[8] The applied procedure considers that each student may have experienced a different academic pathway for the selection of optional exams. The proposed aggregation system succeeds in preserving all the shades of the paths without generating an excessive loss of observations (see section 3.2 for more detail).

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 have been 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 has been 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 face the I-Rat and the teammates' evaluation individually. For each year in which TBL was implemented, the intervention dosage consists 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, therefore, use the structural break from lecture-based to TBL-based course of the same course and the same instructor to evaluate TBL impact on students' achievements controlling for a set of variables that includes also the students' cohort.

Figure 1 – details on Treatment and ITT.



Source: self-elaboration on primary data-
Tools lucid Chart [pie charts are excel imported]

Figure1 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 positive marks only) for females who had not experienced TBL is lower than the male counterpart. Vice versa females who attend TBL succeed 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.

Figure 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 STATA/BE 17.0 in which:

i) a *regression model* has been 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).
ii) a *probit model* has been implemented to find out the probability of passing the macroeconomics exam depending on the characteristics of the students and the academic pathway taken.
iii) a *Cragg's model* (two-part, Hurdle model) has been 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 a likelihoods log test.

All these steps will be developed in detail in paragraph 4.4 dedicated to the econometric analysis.[9]

The methodological approach was organised as follows: initially, intensive data merging and data cleaning work were carried out: several databases of primary data (own and third-party collections) were obtained[10] and merged with administrative data. To summarise briefly data collection also involved the Introductory macroeconomics instructor, current and past Introductory Macroeconomics tutors, the teaching coordinator of the Department of Economics Marco Biagi and the UNIMORE responsible for the administrative data sets.
The set of variables (including TBL experience) whose impact on students' achievements has been shown in the literature has been taken into account and included in the estimated models considering their distribution, the presence of systematic missing for certain groups[11], over time homogeneity[12] and whether their inclusion would impose an excessive reduction of the sample together with their significance in explaining the gender gap in the outcomes.[13] Some variables were excluded from the final model because they do not have significant coefficients and do not contribute to increasing the adjusted-R^2 (*NUTS1 dummies, EnrollGap*, etc..). On the other hand, variables such as *NearbyHighSchool, MathAbility, LogicAbility* and *LowIncome* were retained even though their coefficients were not significant as they contribute to increase the fit of the model.
The selection criterion for the final model was based on a comparison of the following requirements: ① having as greater adjusted R^2 as possible, ② significance of the coefficients (β) and ③ limitation of missing values. If some of them conflict, the adjusted R^2 has been privileged, if the submodels conflict the fit of the female-student model estimate was favoured.

[9] STATA/BE 17.0 has been used to carry out the estimation.
[10] Variables used to generate *IratScore, Dosage* (and *Treat*) originate from six different year databases of primary data as a result of downloads from the Moodle platform. The process repeats each year that TBL was implemented for a total 18 *subdatabase* for the current sample (2021 if we consider the initial sample, which also included the 2020/2021 academic year).
[11] *IraTot* (missing in Treat== 0), *left* and *OverTimeGrad* not reliable for more recent years.
[12] *TOLC; EnglishAbility;etc…*
[13] See Table 2 in section 4.3 and Table A1 in the Appendix for more details.

4.3 descriptive statistic and first findings

Table 2 shows descriptive statistics for the covariates characterizing our sample. Panel A is dedicated to continuous variables, meanwhile, panel B is to dichotomous ones. Both panels consist of two subsections that make gender comparisons between the control (1) and the treated (2) group The typical self-selecting student as treated has a higher average in all subject types (high, middle and Non quantitative); has a short time gap from graduation to university enrolment, and earned more credits in the first year of the undergraduate course attended.

Table 2 – Dependent and Independent Variable Descriptive Statistics

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)
Highlyquantitative	23.36	2.86	23.50	3.10	-0.13	(-0.47)
Slightlyquantitative	22.59	2.94	22.44	2.72	0.15	(0.51)
Nonquantitative	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)
Comprehensionability	5.85	1.88	4.74	2.09	1.11***	(5.27)
Mathability	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)
Gapfromdiploma	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)
N	291		204		495	

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)
highlyquantitative	24.73	3.04	24.77	3.25	-0.03	(-0.11)
slightlyquantitative	24.21	2.97	23.56	3.07	0.65*	(2.00)
nonquantitative	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)
ComprehensionAbility	5.56	2.40	5.06	2.22	0.50*	(2.26)
MathAbility	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	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)
GapfromDiploma	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)
N	294		235		529	

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)
DropOut	0.04	0.20	0.06	0.25	-0.02	(-1.03)
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)
OverTimeGrad	0.38	0.49	0.42	0.49	-0.04	(-0.90)
Winter	0.75	0.43	0.74	0.44	0.02	(0.43)
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)
N	291		204		495	

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)
DropOut	0.02	0.13	0.01	0.11	0.00	(0.40)
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)
OverTimeGrad	0.06	0.24	0.07	0.25	-0.01	(-0.48)
Winter	0.87	0.34	0.84	0.37	0.03	(0.93)
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)
N	294		235		529	

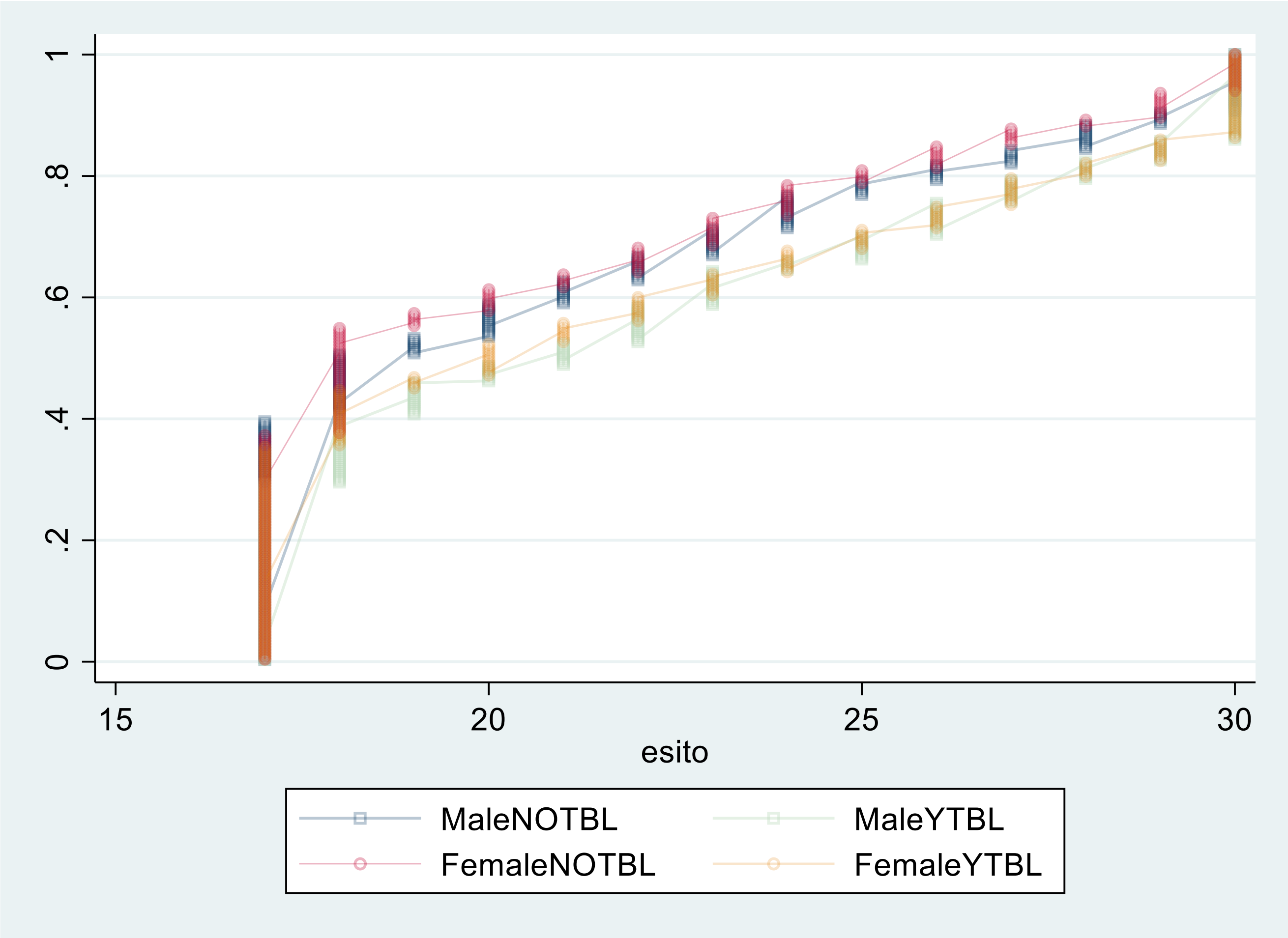
Source: self-elaboration on primary & administrative data.
Notes: t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
[A] TOLC = has changed its composition for students enrolled from 2017 onwards (the English evaluation was introduced).
[B] variable not homogeneous in the sample (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.

This condition is verified for both male and female students and could suggest that students who self-select into treatment are the most deeply motivated. Regardless of the cluster the majority of students have attended a high 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 female students' performance in the entrance test could also be due to the type of question framing. Already previous literature emphasizes how the multiple-choice question type results in disadvantages for females (Reardon et al., 2018) and how the higher risk perceived on average by female students in answering this type of questions impacts on the lower their lower performance (Baldiga, 2014; Karimi & Biria, 2017).

Finally, Figure 2 presents the cumulative distribution of macroeconomics grades of the macroeconomic grades of 4 clusters just presented in table 2.

Figure 2 – Cumulative function of the macroeconomic outcome by gender and TBL attendance.



Source: self-elaboration on primary data
Tool STATA/BE 17.0

What stands out in the table is the truncation that occurs in both the lower and upper bounds: grades have a higher concentration at the edges of the distribution, as all students who score below 18 or above 30 enter the dataset as 17 or 30 respectively. This is why a Hurdle model can be preferred in the estimation to the regression model that does not consider the double truncation in the grades. Another interesting aspect is the left shift of the cumulative function relating to clusters of students who did not experience the TBL. The more the cumulative function is shifted to the left, the more individuals in the group are concentrated in low scores. The figure shows that in the 4 clusters:

- keeping gender constant, having participated in TBL leads to a rightward shift of the curve (better performance)
- keeping the treatment (TBL participation) constant, the male cumulative curve is always to the right of the female one.

4.4 Econometric models

i) Regression model

The estimated model (1) whose results are shown in Table 4 has been obtained after more than 90 trials of models' estimation.

(1) $$Mark_i = \beta_0 + \beta_1 Treat + \beta_2 Female + D_3 Session + \beta_4 Credits + \beta_5 Reteker + \beta_6 Native + \beta_7 LowIncome + \beta_8 Highlyquantitative + \beta_9\, Slightlyquantitative + \beta_{10} nonquantitative + \beta_{11} ComprehensionAbility + \beta_{12}\, MathAbility + \beta_{13} LogicAbility + \varepsilon_i$$

(1a) $$Mark_i\,|_{female} = \beta_0 + \beta_1 Treat + D_2 Session + \beta_3 Credits + \beta_4 Retaker + \beta_5 Native + \beta_6 LowIncome + \beta_7 Highlyquantitative + \beta_8\, Slightlyquantitative + \beta_9 nonquantitative + \beta_{10} ComprehensionAbility + \beta_{11}\, MathAbility + \beta_{12} LogicAbility + \varepsilon_i$$

(1b) $$Mark_i\,|_{male} = \beta_0 + \beta_1 TBLcutoff + D_2 session + \beta_3 credY1 + \beta_4 rep + \beta_5 native + \beta_6 LowIncome + \beta_7 Highlyquantitative + \beta_8\, Slightlyquantitative + \beta_9 nonquantitative + \beta_{10} ComprehensionAbility + \beta_{11}\, MathAbility + \beta_{12} LogicAbility + \varepsilon_i$$

$Mark_i$ is the Macroeconomic performance obtained by a female/male student, 30 is the maximum grade a student can achieve and 18 is the minimum grade that a student can get.
β_1 is the coefficient related to the TBL effect on macroeconomics grades, D_2 regards a set of 16 dichotomous time variables that have been included and account for all unobservable factors that are changing across sessions.
β_7, β_8 and β_9 are coefficients indicating the effects of the student's average[14] exam in the highly quantitative, slightly quantitative and non-quantitative disciplines respectively.
β_{10}, β_{11} and β_{12} return the effect of entry capabilities (before university) i.e. the results of the Entry test in reading comprehension (ComprehensionAbility) in math (MathAbility) and in logic (LogicAbility) on *Mark* and finally ε_i contains all errors resulting from omitted variables and the respective loss of information.
Table 4 illustrates the STATA outcome for models **1, 1a** and **1b** before commenting them, it is important to remember that the regression regarded a dependent variable in which all insufficient marks (which are unobserved) were set to 17.

---

[14] As (due to student's self-selection) each student may have experienced a different academic pathway, subgroups means were computed *ignoring missing* values; for example, if three exams are specified and, some students, select only two, in those observations *newvar* will contain the mean of the two variables that do exist. This procedure makes it possible to minimise the generation of missing values that would otherwise be caused by different academic choices. At the same time it allows a large number of examination marks to be considered. Mean is computed on the mark that have been recorded in the students' academic record reflecting only marks over 17 that have not been rejected by the students after the exam (University of Modena and Reggio Emilia does not include a rule of acceptance of the mark obtained in the exam and it is possible for students to reject the mark obtained and re-take the exam, a maximum of 4 trials of the exams are allowed in an academic year), insufficient marks as well as rejected marks by the students are not detectable.

Table 4 – 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
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The regression in Table 4 shows that participation in TBL is highly significant for all students and exams 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 get almost 3 more points in the exam by taking part in TBL. Being born in Italy seems to be important for male-students but not relevant for female-students' mark in macroeconomics. The positive and significant impact of the *Highlyquantitative, Slightlyquantitative and Nonquantitative* variables coefficients are expected as the level of preparation and ability of the students is linked to all past exams' performance however the impact 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 can be linked to a strategy that is more frequent for female students to sit for the exam as an Attempts to acquire familiarity with the exam structure and then accept only the highest mark and would require higher 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).

ii) Probit model

(1) *Binary outcome of the dependent variable Pass* $\begin{cases} 0 \ if \ student \ fail \\ 1 \ if \ student \ pass \end{cases}$

$$\text{(2)}\ P(pass = 1|X) = \beta_0 + \beta_1 Treat + \beta_2 Female + D_3 Session + \beta_4 Credits + \beta_5 Reteker + \beta_6 Native + \beta_7 LowIncome + \beta_8 Highlyquantitative + \beta_9\ Slightlyquantitative + \beta_{10} nonquantitative + \beta_{11} ComprehensionAbility + \beta_{12}\ MathAbility + \beta_{13} LogicAbility + \varepsilon_i$$

$$\text{(2a)} P\left(pass = 1\middle|X_{female}\right) = \beta_0 + \beta_1 Treat + D_2 Session + \beta_3 Credits + \beta_4 Retaker + \beta_5 Native + \beta_6 LowIncome + \beta_7 Highlyquantitative + \beta_8\ Slightlyquantitative + \beta_9 nonquantitative + \beta_{10} ComprehensionAbility + \beta_{11}\ MathAbility + \beta_{12} LogicAbility + \varepsilon_i$$

$$\text{(2b)}\ P(pass = 1|X_{male}) = \beta_0 + \beta_1 Treat + D_2 Session + \beta_3 Credits + \beta_4 Retaker + \beta_5 Native + \beta_6 LowIncome + \beta_7 Highlyquantitative + \beta_8\ Slightlyquantitative + \beta_9 nonquantitative + \beta_{10} ComprehensionAbility + \beta_{11}\ MathAbility + \beta_{12} LogicAbility + \varepsilon_i$$

The outcome of probit estimations ***2***, ***2a*** and ***2b*** are displayed in table 5 where marginal effects computed at the means of the variables are displayed. Once again, the diversity of the 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 to the TBL seems to be the only variable determining the probability of males passing. This means that participating to the TBL becomes more important than the males' abilities (Highlyquantitative, Slightlyquantitative, Nonquantitative) and their diligence (number of credits acquired in the first year).

Table 5 –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
_cons	-4.802***	-8.074***	-3.668**
	(-5.08)	(-4.41)	(-2.87)
N	585	243	334
CORRECTLY CLASSIFIED	77,44%	81,48%	78,74%

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)

As anticipated in section 4.2 in this analysis we face a censored sample since the value of the dependent variable is not detectable below (or above) a defined threshold. conveniently, unlike the truncated sample, the censored sample is representative of the population[15] because all observations are included, only the dependent variable suffers losses of information.

In our analysis, the test scores of the students who pass are detectable and range from 18 to 30, while we cannot observe the scores of the students who fail the exam (they could have scored 2 points as 17 and for us, it is only known that they do not get a sufficient evaluation). The model

[15] Macroeconomics students in our case.

has, therefore, a lower limit at 17 and -to increase its accuracy- we decided to also consider an upper limit at 30 as it is not possible to distinguish the different marks within excellences.
Tobin (1958) develops a Tobit model that provides consistent and efficient estimators under this restrictive assumption on the dependent variable.
Eq. 4, following Tobin's specification, shows the dependent variable's processing within the model. The actual value for $MacroMark$ is observed if the latent variable $MacroMark^*$ is between 18 and 30 meanwhile lower limit is observed for the censored from below observations and the upper limit is observed for the censored from above observations.

$$\textbf{\textit{Eq 4}} \quad Mark_i = \begin{cases} Mark_i^* & if \quad 18 \leq Mark_i^* < 30 \\ 17 & if \quad Mark_i^* < 18 \\ 30 & if \quad Mark_i^* \geq 30 \end{cases}$$

$$\textbf{\textit{(3)}} \quad Mark_i = \beta_0 + \beta_1 Treat + \beta_2 Female + D_3 Session + \beta_4 Credits + \beta_5 Reteker + \beta_6 Native + \beta_7 LowIncome + \beta_8 Highlyquantitative + \beta_9 Slightlyquantitative + \beta_{10} nonquantitative + \beta_{11} ComprehensionAbility + \beta_{12} MathAbility + \beta_{13} LogicAbility + \varepsilon_i$$

$$\textbf{\textit{(3a)}} \quad Mark_i|_{female} = \beta_0 + \beta_1 Treat + D_2 Session + \beta_3 Credits + \beta_4 Retaker + \beta_5 Native + \beta_6 LowIncome + \beta_7 Highlyquantitative + \beta_8 Slightlyquantitative + \beta_9 nonquantitative + \beta_{10} ComprehensionAbility + \beta_{11} MathAbility + \beta_{12} LogicAbility + \varepsilon_i$$

$$\textbf{\textit{(3b)}} \quad Mark_i|_{male} = \beta_0 + \beta_1 TBLcutoff + D_2 session + \beta_3 credY1 + \beta_4 rep + \beta_5 native + \beta_6 LowIncome + \beta_7 Highlyquantitative + \beta_8 Slightlyquantitative + \beta_9 nonquantitative + \beta_{10} ComprehensionAbility + \beta_{11} MathAbility + \beta_{12} LogicAbility + \varepsilon_i$$

The combination of covariates used to estimate this functional form is the same as in ***reg. 1***, ***1a*** and ***1b*** with the distinction that the dependent variable (*Mark*) is not a simple continuous variable but assume the latent ($Mark$) form in ***Eq 4*** to obtain model ***3***, ***3a*** and ***3b.***

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 1 probability of passing the exam[16] and for the 2 continuous grade variable once a passing grade has been achieved *(Cragg's model* - a Tobit variant*)*.

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

$$\textbf{\textit{Eq 5}} \quad \lambda = 2 * (\mathrm{LL}_{probit} + \mathrm{LL}_{truncreg} - \mathrm{LL}_{tobit})$$

Manual application of the formula is reported in addendum A5 in Appendix 1 where the chi-square test validates the best fit of the Cragg's model. This condition is verified both for the main model and for its gender disaggregation.

[16] Which we have already presented in eq. 2 and respective outputs in tab 5.

Table 6 – Output of models 3, 3a and 3b.

Coefficients are omitted and marginals effects are displayed:
dydx(*) at means predict(e(17,.)), E(Mark| Mark>17), predict(e(17,.))

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

As Cragg's model develops in two consecutive stages, the output of table 6 must be analysed in light of the findings of the probability of being promoted in the Probit model (table 5). Table 6 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 $Mark$[17].
Observing the first column of table 6, we note that -once promoted- having participated in the TBL increases the examination' marks by a 3points, 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 were attended classes when TBL has been 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.

5 CONCLUSION AND FUTURE RESEARCH DEVELOPMENT

Using different cohorts of the Introductory Macroeconomics students, the primary aim of this paper is to investigate the impact of TBL on students' achievement measured by the exam marks and the pass probability controlling for a set of observable variables that have been found in the literature to affect grades.

> **H_1 =** *attending Team-Based Learning Lessons produces better learning outcomes in terms of course exam grade.*

H_1 seems to be confirmed by the positive coefficient of column 1 in table 4 and even more by the more refined estimates produced by the marginal effects of the Hurdle Model. This evidence is also consistent with the literature reviewed in Section 2, which demonstrates that those who take a TBL course are more likely to get higher grades.

> **H_2 =** *students react differently to treatment depending on their gender.*

Also, this second hypothesis seems to be verified. Female students participating in TBL significantly increase their grades, but have no effect on the likelihood of passing the exam. In contrast, male participants have small effects on grades, but a strong impact on the probability of passing the exam. The positive impact on female' grade is again consistent with the literature and can also be related to the impact of TBL structure of tests in letting female students, who have been shown by the literature to be more negatively affected by multiple replies type of questions, to be trained during the course on this type of test by the recurrent tests with this structure, thus acquiring confidence and skills in facing a type of exam otherwise on average more difficult to deal with for them.
Meanwhile the two joint results (impact on grades and probability of passing) might suggest that the practice of TBL not only influenced knowledge but also changed gender behaviour.
The literature shows that females, who are more risk-averse, tend to show up for the exam only if they are fully prepared, while males, in contrast, tend more to "*try it*". So TBL practice on males may have also acted on their approach to examination by empowering them.

> **H_3 =** *Female performance in Macroeconomics is lower.*

---

[17] i) coefficient computed with marginal effect are lower respect those related to the latent variable. ii) Coefficient computed with marginal effect change according to the values given to the covariate.

The assumption that females perform worse in macroeconomics (as a quantitative subject) seems to be rejected by our analyses. In fact, although in descriptive statistics (table 2, panel A) it is shown that women score significantly worse on the entrance test (TOLC and any of its subgroups), they manage to achieve an academic performance in line than that of men (or even higher if they attended TBL). A possible explanation could reside in the structure of the TOLC test: there is a widespread literature which states that female performance is penalized by multiple choice tests (Baldiga, 2014; Karimi & Biria, 2017; Griselda 2020). Supporting this hypothesis is the fact that females also underperform in verbal comprehension in the entry test (TOLC) , whereas they usually tend to be more talented in this latter. This point is another research goal we are trying to achieve by analysing the structural break of the Covid pandemic that caused the massive use of this testing modality.

H_4 = *TBL could help overcome gender differences in macroeconomics.*

Even if our analysis does not show significant gender divergences in macroeconomics exams marks (see **H_3**) participating in TBL has an extremely positive impact on female outcomes (see **H_2**) and this can help to prevent the - now not very significant - gap from widening. In addition, as **H_2** reports , TBL teaching may have had effects not only on the grade itself in economics but also on behaviors (see approach to exam).

Taken together, these results suggest that there is a positive and significant association between attending TBL's courses and macroeconomic exam performance, controlling for a set of individual variables connected both with students' socio-demographics and cognitive skills.

However, we are aware of the limitation 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 for the exam as not attending students. Most of the students who had the option of attending the TBL course did opt in however we could not exclude students' self-selection into treatment though it is conceivable to assume that the covariates included in the analyses 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 Instructor by following a lecture-based approach without any opportunity to have TBL sessions, to improve the evaluation of the impact of TBL. Alternatively and/or additionally, it is suggested to include the Heckman (1979) correction for non-random selection in the treatment. Finally, application of the Oaxaca decomposition (Blinder 1973; Oaxaca 1973) proposed by Bauer and Sinning (2010)[18], can allow to detect the gender differential in Introductory Macroeconomics not due to differences in the observed characteristics.

Further developments will include evaluation of other likely TBL outcomes as observed in the literature as level of engagement in the learning process and development of problem solving, teamwork skills as well as more collaborative behaviour and degree of attractiveness of TBL courses for disadvantaged groups of the students' population. And, last but not least, 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 inspired at Climate Survey on Diversity, Equity and Inclusion.

[18] 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 & Sinning, 2010).

Appendix 1

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 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 is 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 Attemptsing the exam for the first time. ******variable obtained by manipulating Attempts.*
PERFORMANCE IN HIGHLY QUANTITATIVE SUBJECT	*Highlyquantitative*	Continuous variable which computes the mean[A] of students on exams which have a high quantitative content. By analysing the university courses offering by the department of Economics and filtering out the exams not taken by any of the students in the sample, a pool of 14 exams was created having a highly quantitative content *[see Table A2 for more details]* *[A] mean is computed on the verbalised mark, insufficient marks as well as rejected marks are not detectable.* **Row subgroups means were computed ignoring missing values in the pool.*
PERFORMANCE IN SLIGHTLY QUANTITATIVE SUBJECT	*Slightlyquantitative*	Continuous variable which computes the mean[A] of students on exams which have a medium quantitative content. By analysing the educational offer of the *department* of economics and filtering out the exams not taken by any of the students in the sample, a pool of 7 exams with a slightly lower quantitative content has been detected *[see Table A2 for more details]* *[A] mean is computed on the verbalised mark, insufficient marks, as well as rejected marks, are not detectable.* **Row subgroups means were computed ignoring missing values in the pool.*
PERFORMANCE IN NON QUANTITATIVE SUBJECT	*Nonquantitative*	Continuous variable which computes the mean[A] of students on exams which have not a quantitative content. By analysing the educational offer of the department of economics and filtering out the exams not taken by any of the students in the sample, a pool of 22 exams not having a quantitative content hase been selected *[see Table A2 for more details]* *[A] mean is computed on the verbalised mark, insufficient marks, as well as rejected marks, are not detectable.* **Row subgroups means were computed ignoring missing values in the pool.*
NEIGHBOURHOOD	*NearbyHighSchool*	Dummy variable which has a value of 1 if the student attended a high school in the same region as the university and 0 otherwise. **We are aware that the accuracy of the variable is weak for neighbouring regions (as there may be municipalities in other regions closer than those in Emilia Romagna itself), our intention in the medium term is to adjust for the proximity of municipalities even if they are not located in the region.*
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 MATH	*MathAbility*	Continuous variable which reports students' performance in math 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 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 was not missing. **variables used to generate Dosage originate from six different databases of primary data as a result of downloads from the Moodle platform. The process repeats each year TBL was implemented for a total 18 subdatabase for this sample.*
PARTECIPATION AT TBL	*Participation*	Dummy variable equal to 1 if student participated at least at one Team-Based Learning' lessons and 0 otherwise. *[uneven and fragmented participation rate]* **variables used to generate Participation originate from six different databases of primary data as a result of downloads from the Moodle platform. The process repeats each year TBL was implemented for a total 18 subdatabase for this sample.*
INDIVIDUAL PERFORMANCE AT I-RAT	*IratScore*	Continuous variable which computes the mean of all Irat score collected by students. **variables used to generate IratScore originate from six different databases of primary data as a result of downloads from the Moodle platform. The process repeats each year TBL was implemented for a total 18 subdatabase for this sample.*
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). **variable generated through lags using session as units of time.*
WINTER SESSION	*Winter*	Dummy variable has a value of 1 if the student takes the exam in the January or February sessions (which are closest to the semester when the macroeconomics course is taught) and 0 if they wait for Summer sessions. **Not in the principal analysis because of collinearity with dummy sessions and because we are aware that is not neutral to TBL (students who participate to TBL tends to cluster themselves in winter session)*
ECONOMIC HARDSHIP INDICATOR	*LowIncome*	Dummy variable equal to 1 if student's family unit has an equivalent economic status indicator lower than 23.000 € and 0 otherwise.
ECONOMIC HARDSHIP INDICATOR	*MiddleIncome*	Dummy variable equal to 1 if student's family unit has an equivalent economic status indicator lower than 45.000 € and 0 otherwise.
ECONOMIC HARDSHIP INDICATOR	*HighIncome*	Dummy variable equal to 1 if the student has not applied for fee reductions. This suggests that student's family unit has an equivalent economic status indicator higher than 45.000 € and 0 otherwise.
TEST ONLINE CISIA	*Tolc*	Continuous variable that reports the students' university entry tests results in which students must solve math, logic and reading comprehension questions. The result of each individual test is determined by the number of correct, wrong and not given answers that determine an absolute score, deriving from 1 point for each correct answer, 0 points for each answer not given and a penalty of 0.25 points for each wrong answer. **Not in the principal analysis because it changes across year by the introduction of the English test sub-questions for students who have enrolled from 2017 onwards.*
COMPLIANCE AT THE END OF THE LAST YEAR	*OverTimeGrad*	Dummy variable equal to 1 if the student has not graduated within the prescribed period (April+1 in the last academic year) and 0 otherwise. **Our reference sample does not allow us to use this variable because they are not all students who should already have graduated (the last cohorts of the analysis are students who, in the A.Y. 2019/2020, are in their second*

		year and should have graduated by April 2022. Furthermore, we are not sure to have the updated data even for the A.Y. 2018/2019).
COMPLIANCE AT THE END OF THE FIRST YEAR	*Credits*	Dummy variable equal to 1 if the student gets at least 40 credits in the first year and 0 otherwise.
DO NOT COMPLETE UNIVERSITY	*DropOut*	Dummy variable equal to 1 if the student left the university without graduating. **As above, but here the time constraint is more relaxed and we expected even less accuracy of the* ***OverTimeGrad*** *variable.*
ENROLMENT WAITING PERIOD	*EnrollGap*	Continuous variable that corresponds to years from graduation to university enrolment.
NEIGHBOURHOOD (NUTS1 ARRANGEMENTS)	*Northeast* *Northwest* *Center* *South&Islands*	4 Dummy variable which takes the value of 1 if the student attended a high school in one of those macroareas (NUTS1) and 0 for the others.

Table A2 – Disaggregation of macro-structures of examination performance

NAME	subject	class
TeachingActivityEC_EIF01	Economics of financial intermediaries	*Highly Quantitative*
TeachingActivityEC_EM01	Monetary economics	*Highly Quantitative*
TeachingActivityEC_FA01	Corporate finance, financial analysis	*Highly Quantitative*
TeachingActivityEC_FA02	Corporate finance	*Highly Quantitative*
TeachingActivityEC_IMI01	Introduction to microeconomics	*Highly Quantitative*
TeachingActivityEC_MA02	Macroeconomics	*Highly Quantitative*
TeachingActivityEC_MEF01	Mathematics for economics and finance	*Highly Quantitative*
TeachingActivityEC_MFA01	Financial and actuarial mathematics	*Highly Quantitative*
TeachingActivityEC_MI03	Microeconomics	*Highly Quantitative*
TeachingActivityEC_MIF01	Models for financial investments	*Highly Quantitative*
TeachingActivityEC_MMF01	Mathematics and financial mathematics	*Highly Quantitative*
TeachingActivityEC_RSFF01	Savings and financial choices of enterprises	*Highly Quantitative*
TeachingActivityEC_SF01	Financial science	*Highly Quantitative*
TeachingActivityEC_ST01	Statistics	*Highly Quantitative*
TeachingActivityEC_EA01	Business economics	*Slightly Quantitative*
TeachingActivityEC_EA02	Business economics 2	*Slightly Quantitative*

TeachingActivityEC_EAC01	Economics of credit companies	*Slightly Quantitative*
TeachingActivityEC_EAC02	Economics of credit companies	*Slightly Quantitative*
TeachingActivityEC_EMM01	Securities market economics	*Slightly Quantitative*
TeachingActivityEC_SW01	Welfare systems	*Slightly Quantitative*
TeachingActivityEC_MA01	Marketing	*Non-quantitative*
TeachingActivityEC_DI01	Industrial law	*Non-quantitative*
TeachingActivityEC_DL02	Labour law	*Non-quantitative*
TeachingActivityEC_DP01	Public law	*Non-quantitative*
TeachingActivityEC_DPC01	Private and commercial law	*Non-quantitative*
TeachingActivityEC_DT01	Tax law	*Non-quantitative*
TeachingActivityEC_DUE01	European Union law	*Non-quantitative*
TeachingActivityEC_EGI01	Economics and business management	*Non-quantitative*
TeachingActivityEC_EI01	International economics	*Non-quantitative*
TeachingActivityEC_EIDI01	Economics and institutions of industrial districts	*Non-quantitative*
TeachingActivityEC_EPL01	Economics and labour policies	*Non-quantitative*
TeachingActivityEC_ERS01	Ethics and corporate social responsibility	*Non-quantitative*
TeachingActivityEC_IEPC01	EU integration and community policies	*Non-quantitative*
TeachingActivityEC_MI01	International marketing I	*Non-quantitative*
TeachingActivityEC_MI02	International marketing II	*Non-quantitative*
TeachingActivityEC_OA01	Business organisation	*Non-quantitative*
TeachingActivityEC_PC01	Programming and control	*Non-quantitative*
TeachingActivityEC_RM01	Marketing research	*Non-quantitative*
TeachingActivityEC_SE01	Economic history	*Non-quantitative*
TeachingActivityEC_SEI	Italian economic history	*Non-quantitative*
TeachingActivityEC_SPE01	Economic history	*Non-quantitative*
TeachingActivityEC_SR	Social responsibility	*Non-quantitative*

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	-0.13	-0.01	0.14^{*}
IratScore	-0.04	-0.14^{*}	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)**).

Table A4 – Goodness of the probit fitting

GLOBAL MODEL

```
              -------- True --------
Classified |         D            ~D  |      Total
-----------+--------------------------+-----------
     +     |       406           102  |        508
     -     |        30            47  |         77
-----------+--------------------------+-----------
   Total   |       436           149  |        585

Classified + if predicted Pr(D) >= .5
True D defined as DUMesito != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   93.12%
Specificity                     Pr( -|~D)   31.54%
Positive predictive value       Pr( D| +)   79.92%
Negative predictive value       Pr(~D| -)   61.04%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   68.46%
False - rate for true D         Pr( -| D)    6.88%
False + rate for classified +   Pr(~D| +)   20.08%
False - rate for classified -   Pr( D| -)   38.96%
--------------------------------------------------
Correctly classified                        77.44%
```

FEMALE MODEL

```
              -------- True --------
Classified |         D            ~D  |      Total
-----------+--------------------------+-----------
     +     |       162            28  |        190
     -     |        17            36  |         53
-----------+--------------------------+-----------
   Total   |       179            64  |        243

Classified + if predicted Pr(D) >= .5
True D defined as DUMesito != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   90.50%
Specificity                     Pr( -|~D)   56.25%
Positive predictive value       Pr( D| +)   85.26%
Negative predictive value       Pr(~D| -)   67.92%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   43.75%
False - rate for true D         Pr( -| D)    9.50%
False + rate for classified +   Pr(~D| +)   14.74%
False - rate for classified -   Pr( D| -)   32.08%
--------------------------------------------------
Correctlv classified                        81.48%
```

MALE MODEL

```
              -------- True --------
Classified |         D            ~D  |      Total
-----------+--------------------------+-----------
     +     |       230            52  |        282
     -     |        19            33  |         52
-----------+--------------------------+-----------
   Total   |       249            85  |        334

Classified + if predicted Pr(D) >= .5
True D defined as DUMesito != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   92.37%
Specificity                     Pr( -|~D)   38.82%
Positive predictive value       Pr( D| +)   81.56%
Negative predictive value       Pr(~D| -)   63.46%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   61.18%
False - rate for true D         Pr( -| D)    7.63%
False + rate for classified +   Pr(~D| +)   18.44%
False - rate for classified -   Pr( D| -)   36.54%
--------------------------------------------------
Correctly classified                        78.74%
```

Addendum A5 - Fit test: Simple Tobit VS Craggs Model

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

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

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

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.

6 REFERENCES

Abio, G., M. Alcaniz, M. Gomez-Puig, G. Rubert, M. Serrano, A. Stoyanova, and M. Vilalta-Bufi. (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.

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

Ballester, Pilar, C. '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, & Mathias Sinning. «Blinder-Oaxaca decomposition for Tobit models». *Applied Economics* 42, n. 12 (2010): 1569–75. https://doi.org/10.1080/00036840701721612

Bellini, C.; De Santis, A.; K. Sannicandro, K.; Cecconi, L. (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, 2020

Blinder, Alan S. 'Wage Discrimination: Reduced Form and Structural Estimates'. *Journal of Human Resources* 8, no. 4 (1973): 436–55.

Bilen, E., & Matros, A. (2021). Online cheating amid COVID-19. Journal of Economic Behavior & Organization 182, 196-211. doi:10.1016/j.jebo.2020.12.004.

Burke, William J. (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, G. & Ghanei, M. (2022) 'Team-Based Learning in Economics: Promoting Group Collaboration, Diversity and Inclusion', *Journal of Economic Education*, 53(1), pp. 11–30.

Castagnetti, Carolina, and Luisa Rosti. (2009). 'Effort Allocation in Tournaments: The Effect of Gender on Academic Performance in Italian Universities'. *Economics of Education Review* 28 (3): 357–69. https://doi.org/10.1016/j.econedurev.2008.06.004.

Coyne, Leanne, Jody K. Takemoto, Brittany L. Parmentier, Thayer Merritt, and Rachel A. Sharpton. (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.

Clerici-Arias, M. (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.

De Santis, A.; K. Sannicandro, K.; Bellini, C.; Cecconi, L. (2019). Valutazione e soft skills nella didattica universitaria, Training actions and evaluation processes, Atti del Convegno Internazionale SIRD, pp. 493-502.

Espey, M. (2022) 'Variation in Individual Engagement in Team-Based Learning and Final Exam Performance', *International Review of Economics Education*, 41.

Espey, M. (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'. https://doi.org/10.22004/AG.ECON.123521.

Griselda, S. (2020). Different Questions, Different Gender Gap: Can the Format of Questions Explain the Gender Gap in Mathematics? Job Market Papers pgr710.

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 (April): 119884. https://doi.org/10.1016/j.jclepro.2019.119884.

Good, C., Rattan, A., & Dweck, C. S. (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, J. (1979). 'Sample Selection Bias as a Specification Error', *Econometrica*, Vol. 47, No. 1, 153–161.

Hettler, P. L. (2015). Student demographics and the impact of team-based learning. *International Advances in Economic Research* 21 (4): 413–22.

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

Jakobsen, K.V., & Knetemann, M. (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, M., & Biria, R. (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. doi:10.17507/tpls.0710.10.

Michaelsen, L., Neil Davidson, and C. Major. (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, L. K., and Michael Sweet. (2008). 'The Essential Elements of Team-Based Learning'. *New Directions for Teaching and Learning* 2008 (116): 7–27. https://doi.org/10.1002/tl.330.

Michaelsen, L.K., Knight, A.B., & Fink, L.D. (2004a). Centers for Teaching and Technology. Available at https://bit.ly/3sJzYSQ, last accessed on 22nd April 2021.

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

Odell, Kathleen E. 2018. ‘Team-Based Learning and Student Performance: Preliminary Evidence from a Principles of Macroeconomics Classroom’. *International Review of Economics Education* 29 (September): 44–58. https://doi.org/10.1016/j.iree.2018.01.001.

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

Parappilly, M., Woodman, R.J. & Randhawa, S. (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, D.X., & Michaelsen, L.K. (2010). Twelve tips for doing effective Team-Based Learning (TBL). Medical Teacher 32 (2), 118-122. doi:10.3109/01421590903548562.

Reardon, S.F., Kalogrides, D., Fahle, E.M., Podolsky, A., & Zárate, R. (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). doi:10.3102/0013189X18762105.

Ruder, P., Maier, M.H. & Simkins, S.P. (2021) ‘Getting started with team-based learning (TBL): An introduction’, Journal of Economic Education, 52(3), pp. 220–230.

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.

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

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

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

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

Acknowledgments.

We thank the Introductory Macroeconomic Course instructor and tutors, the teaching coordinator of the Department of Economics Marco Biagi and the UNIMORE 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 a Ph.d. Seminar in the Labour, Development and Innovation Phd series of seminars for their insightful comments. The responsibility for the content of the paper and any errors remains exclusively with the authors.