education sciences

MDPI

Article

Analysis of the Effectiveness of Open-Ended Style Activities on Students' Performance in an Engineering Admission Test

Roberto Virzi *, Matteo Bozzi, Marco Costigliolo, Roberto Luca Mazzola and Maurizio Zani

Department of Physics, Politecnico di Milano, 20133 Milano, Italy; matteo.bozzi@polimi.it (M.B.); marco.costigliolo@polimi.it (M.C.); roberto.mazzola@polimi.it (R.L.M.); maurizio.zani@polimi.it (M.Z.)

* Correspondence: roberto.virzi@polimi.it

Abstract

In the academic year 2022/2023, an orientation course addressed to high school students was proposed at Politecnico di Milano. The course was conducted using active methodologies referring to the Problem-Based Learning pedagogical framework. Most of the time was dedicated to an open-ended style physics laboratory in which students worked in small groups exploring different areas of physics. The main aim of the course was to foster science self-efficacy and motivation, in order to enable students to prepare themselves for passing the test required to enrol in any engineering programme at Politecnico di Milano University. To investigate the effectiveness of the course, a statistical analysis of students' attendance and students' results on the test was performed. The results of the analysis show that students who attended the course had a slightly better improvement in their test scores compared to those who did not attend. The impact of the course seems to be more effective for female students. The results confirm the validity of active and open-ended activities to increase scholastic performance and to enable students in autonomous preparation. Using these strategies in an orientation course can help students make more informed choices about the university pathway best suited to them, thereby reducing issues related to student dropout.

Keywords: PBL; open-ended; orientation course; effectiveness; self-efficacy; engineering test

1. Introduction

In schools and universities, teachers are constantly interested in finding new strategies to improve students' learning experience. Nowadays, teachers can rely on a wide set of different methodologies like student-centred approaches (Bozzi et al., 2021, 2023; De Backer et al., 2015; Edström & Kolmos, 2014; Hattie, 2023; Jones, 2007; Vickrey et al., 2015; Wood, 2023) or technology-enhanced learning (Bozkurt, 2020; Istenič, 2024; Kirkwood & Price, 2014; Lam et al., 2021; Otto et al., 2024), but a fundamental thing to consider is their actual effectiveness. Considering a specific context, an educator could aim to improve the motivation of their pupils or their abilities in particular topics or areas. For instance, concerning the teaching of physics, we are certainly interested in making our students aware of the scientific method they have to apply and be autonomous in making correct reasoning and interpretation. Traditional physics teaching consists of frontal explanations, guided solving of exercises and guided laboratory experiences, which seem to have low efficacy (Hake, 1998). In particular, in traditional physics laboratories, students are asked to follow precise instructions in order to replicate a specific experiment or confirm the validity

Academic Editor: Andrea C. Borowczak

Received: 9 February 2026
Revised: 17 March 2026
Accepted: 19 March 2026
Published: 21 March 2026

 https://doi.org/10.3390/educsci16030489

of a particular law. For this reason, these kinds of laboratories are also referred to as the "cookbook approach" (Doucette et al., 2018; Holmes & Wieman, 2018).

On the contrary, innovative techniques involve student-centred tasks where each learner is asked to be the protagonist of the activity to achieve learning goals better (Smith & Holmes, 2021). This means that students should discuss with peers the right procedure to follow or how to set up an experiment. In this direction, the do-it-yourself style (Craig & Hill, 2012) requires students to design an experiment on their own to measure a certain quantity and understand a phenomenon. This methodology is always present in open-ended (OE) style laboratories (Trumper, 2003). In a laboratory session performed in an OE style, students are asked to investigate a phenomenon with few indications and little advice. Previous studies highlighted that this innovative strategy helps students reinforce their science self-efficacy and their interest in scientific topics (Alhadabi, 2021; Bandura, 1997; Wilcox & Lewandowski, 2016). This teaching strategy is an application of the Problem-Based Learning (PBL) approach (Celik et al., 2011). PBL is a student-centred teaching method in which learning starts with a real-world problem rather than a lecture or a list of facts. Instead of being told the solution, students work in small groups to analyse the problem, identify what they need to learn, research information, and propose solutions.

The teacher acts as a facilitator or guide, helping students reflect, ask good questions, and connect their findings. PBL is grounded in constructivism theory, which holds that learning takes place through the active construction of knowledge in the learner's mind (Bodner, 1986).

In another study, an OE strategy was also used to evaluate students' comprehension of a common physics phenomenon (Doğru & Kurnaz, 2023). In this last example, thanks to an open-ended questionnaire, it has been revealed that students involved had difficulties in providing a correct explanation of physics phenomena, even though these are related to a common daily experience. This shows that in that situation, students probably did not have a full and deep comprehension of these phenomena.

As stated before, an important step is the evaluation of the actual effectiveness of an innovative learning strategy. In most cases, researchers did this evaluation by taking advantage of some preexisting physics tests like the Force Concept Inventory (FCI) (Hestenes et al., 1992). For instance, in (Hoellwarth et al., 2005; Redish et al., 1997), the effectiveness was obtained by the gain in a pre-post submission of the FCI to the students in groups that attended traditional lectures and groups that participated in innovative lectures. In other studies (Gök, 2011), researchers also focused on the improvements concerning motivation and self-efficacy by submitting the Motivated Strategies for Learning Questionnaire (Pintrich & Others, 1991). It is interesting to notice that in this last example, the authors highlighted that an improvement in comprehension of phenomena is not necessarily related to an improvement in problem-solving skills and vice versa. In (Doucette et al., 2018; Holmes & Wieman, 2018), an analysis of the efficacy of innovative laboratories has been conducted, highlighting positive effects, especially in motivation and critical thinking.

In the past, our group has also been involved in other research related to the transition from high schools to STEM (Science, Technology, Engineering, Mathematics) universities. In particular, we investigated the validity of using Massive Open Online Courses (MOOCs) to improve the teaching of physics and support students in choosing their future university path (Raffaghelli et al., 2018). We also took into consideration the positive effect of a refresher course for freshmen (Gondoni et al., 2021) and of a physics laboratory conducted among a set of peculiar activities in Italian high schools named Percorsi per le Competenze Trasversali e l'Orientamento [Programs for General Skills and Orientation] (Mazzola et al., 2021).

https://doi.org/10.3390/educsci16030489

In this work, we were interested in evaluating the effectiveness of our OE activity in terms of how our students, after attending our course, improved their results in the Test On-Line (TOL) required to enter an engineering programme at Politecnico di Milano. Unfortunately, because of some limitations that will be described in further sections, we were not able to find a causality relationship between the attendance of the course and an improvement in the TOL. Nevertheless, we could observe an interesting tendency. Before starting our analysis, it is important to take some aspects of the course into consideration.

The TOL is a computer-based multiple-choice test composed of 65 questions, which are organised as follows: 30 questions about English, 25 questions about Mathematics, 5 questions about Physics, and 5 questions about Verbal Comprehension. Students get 1 point for each correct answer, 0 for a missing one and a negative 0.25 for a wrong choice. Each section has a different weight, so the final score ranges from −25 to 100 points (*How to access to Laurea programmes in engineering*, 2025) as reported in Table 1.

Table 1. Resume of TOL structure.

Topic	No. of Items	Weight	Min Score	Max Score
English	30	1/3	−2.5	10
Mathematics and Logic	25	2.6	−16.25	65
Physics	5	2	−2.5	10
Verbal Comprehension	5	3	−3.75	15
Total	65		−25	100

Traditionally, a course dedicated to the preparation for an engineering test consists of a significant amount of hours of frontal lectures or guided solving of exercises and mock tests. In our situation, we could rely on only 15 h to cover all the activities and, more specifically, on only 6 h for the physics part. Hence, we organised our time in an innovative style to obtain the best possible results. Our main aim could not be to provide students with a detailed review of all the topics covered in the TOL, nor to show them how to solve a proper number of standard exercises. On the contrary, we aimed to make students autonomous in their personal preparation for the test. To reach this goal, we chose to dedicate most of the time to an OE laboratory session, focusing our attention on both the extensive use of feedback (Hattie & Clarke, 2018) and the promotion of individual self-efficacy. The main idea is that, following these strategies, students can improve their performance in the TOL thanks to the enhancement of their learning skills, even without a traditional course.

Indeed, according to (Marušić & Sliško, 2014), an active approach has a positive effect on students in terms of motivation and self-efficacy. This happens primarily because students experience positive feelings during the activities, they find the topic more attractive, and they are encouraged in critical thinking. Similarly, in (Celik et al., 2011), it is shown that the use of a PBL strategy is correlated with a better achievement of physics topics. This is probably because PBL methodology is centred on the students and helps to construct the information to obtain meaningful learning. In (Gök, 2011), it is reported that students with a higher self-efficacy had a better scholastic performance. The reason for that is probably because self-efficacy fosters the student's perseverance in obtaining specific results and improves their willingness to deal with demanding tasks. Eventually, in another study (Alhadabi, 2021), a correlation between gender and the impact of science self-efficacy on scholastic results was found. More precisely, strengthening science self-efficacy seems to be more effective on female students, especially those with low socio-economic status. Moreover, it has been observed that the low representation of women in scientific university programmes is mostly due to psychological factors like a low science self-efficacy (Jungert et al., 2019; Marsh et al., 2019). For this reason, working on fostering self-efficacy and motivation should be useful, especially for female students.

These considerations lead to the following research questions:

RQ1: Is our course actually effective in improving TOL results for students?

RQ2: Are open-ended activities actually helpful for the preparation of the TOL?

RQ3: Is the gender of students correlated to the effectiveness of our course?

2. Materials and Methods

2.1. Course Structure and Research Design

In the academic year 2022/2023, our research group ST2 ran a part of an orientation course addressed to high school students interested in preparing themselves for passing the TOL to enter an engineering programme at Politecnico di Milano. The course was named "Introduzione metodologica alla preparazione al test di Ingegneria" (Methodological Introduction to Engineering Test Preparation), and it has also been proposed for the following academic years. The organisation of this course has been made possible thanks to funding related to PNRR (National Recovery and Resilience Plan) (*DM 934/2022*, 2022). Any local high school could present the course to its students, who were free to enrol. To better organise the activity, several editions were scheduled, each limited to 25–30 students from the same school. Each session was therefore dedicated to a single school.

The course was structured in 15 h, divided into 3 main groups. A total of 3 h were dedicated to a frontal lecture about motivation, 6 h were dedicated to mathematics activities and the last 6 h were dedicated to physics activities. Our group managed this last part by proposing 2 h of an introductory online meeting about the study method for physics and 4 h of OE physics laboratory. During the online meeting, students answered a propaedeutical physics questionnaire validated in previous works of our research group (Virzi et al., 2025a, 2025c). During the laboratory, students were divided into small groups (3–5 students per group), and each group received a few general indications about the phenomenon they were asked to study and a box of fundamental materials. Students had to discuss and decide how to set up the experiment and which quantities were useful to measure. They were encouraged to find possible dependence or independence factors between quantities. In the end, each group had to report their results to the rest of the class and answer possible questions from classmates or tutors. Each group worked on a different topic/phenomenon among the following: free motion and friction; gravity and pendulum; springs and oscillatory motions; optics; volume of irregular objects; motion in a viscous medium; and conservation of mechanical energy.

Each group was also asked to take note of their decision, their data and their results. This was useful not just to better organise their final presentation but also to make students aware of the importance of collecting, storing and sharing data with peers in a scientific context. We decided to underline this concept because this laboratory session has also been presented as a small experience of what it means to work like a physics researcher.

Both during the lab and at the end of the activity, students received feedback from the instructors. In particular, tutors circulated among the groups, asking about the progress of their work and providing hints when necessary. During the final presentations, tutors asked each group to explain why they had chosen a particular strategy and provided feedback on the correctness of their procedures. The main aim was to promote discussion among peers; however, only a few questions or interventions came from the students. In this first edition, 113 students participated in the activity, spread across 5 consecutive editions between March and April 2023.

We wanted to evaluate the effectiveness of our activities by observing the results of our students in the TOL sustained in the same academic year. According to what we exposed in Section 1, we believe that an OE activity could help students become able to prepare themselves autonomously for the TOL. We expected to observe an improvement in TOL results before and after the attendance of the course.

2.2. Groups of Students

To properly answer the research questions, we had to build two groups of students starting from the whole database of students who tried the TOL in academic year 2022/2023. The first (G1) is composed of students who attended our course and tried the TOL. The second (G2) contains students who did not participate in our activities. We wanted to look for improvements, so we needed students who had tried the TOL more than once. This is not an uncommon situation: students are allowed to try the test several times if they are not satisfied with their results. Indeed, a high final score is important to increase the probability of being admitted to the desired engineering programme. This means that candidates who had more than one trial wanted to improve their scores. Evidence of that is the fact that in the academic year 2022/2023, the TOL was tried 21,606 times by 13,412 different persons. Among them, only 7550 students (56% of the total) had a single attempt.

To build the G1, we had to consider students who had tried the TOL at least one time before the attendance of the course and at least one time after the attendance. Among the 113 students who attended the course, we found only 25 with these characteristics.

To build the G2, we selected the students who sustained the TOL at least twice. In this case, these trials are obviously not related to when our course was developed. To properly compare this sample with the G1, we decided to select only students belonging to the same high school as the students included in the G1. This option allowed us to state that the two groups are homogeneous enough, as we discuss in the next subsection. Following this rule, we built a G2 composed of 99 students.

To be more precise, G1 is composed of 18 male students and 7 female students. According to the Italian high school organisation, 15 of them came from a scientific high school (Liceo Scientifico), 8 from a technical high school (Istituto Tecnico Industriale) and 2 from a humanities-oriented high school (Liceo Classico). Further, 16 were in their 4th year of high school, and 9 were in the 5th (and last) year. As far as G2 concerns, 58 were male students and 41 were female. A total of 74 came from a scientific high school, 5 from a technical high school and 19 from a humanities-oriented high school. Finally, 55 were in their 4th year of high school and 44 were in the 5th (and last) year. All this information is summarised in Table 2.

Table 2. Composition of G1 and G2.

	G1	**G2**
Total number	25	99
Male	18	58
Female	7	41
Scientific school	15	74
Technical school	8	5
Humanities-oriented school	2	19
4th year	16	55
5th year (final)	9	44

Because students in G1 and G2 were not selected randomly, G2 cannot be considered as a control group referred to as study group G1. For this reason, we only used these two groups to make a comparison between students who attended our course and students who did not attend, using G2 as a baseline. The absence of a control group is surely a limitation to our research; nevertheless, some interesting results will be presented in Section 4. The dataset with all information about G1 and G2 is available on the Zenodo website (Virzi et al., 2025b). It is interesting to report that, among the 113 participants in the course, 10 decided not to try the TOL. We will give a comment about this in the Discussion section.

2.3. Homogeneity of Groups

To ensure that our analysis is trustworthy, we wanted to check the homogeneity of G1 and G2. The first attention we had was to choose students from the same high schools as stated in the previous subsection. In addition, we investigated the results in the INVALSI test (www.invalsi.it, accessed on 3 March 2025), the Eduscopio ranking (www.eduscopio.it, accessed on 3 March 2025) and the socio-economic situation of the schools included in the samples related to academic year 2022/2023.

INVALSI is the acronym of "Istituto Nazionale per la VALutazione del Sistema educativo di Istruzione e di formazione" (National Institute for the Evaluation of the Education and Training System) and consists of the administering of a test about Italian language and Mathematics to students in different school years. More precisely, the INVALSI test is administered to students in the second and fifth years of primary school, the third year of middle school, and the second and fifth years of high school. The aim of this evaluation protocol is to provide a precise indication of the level reached by students at some point in their educational path. The INVALSI results are released sorted by region, province, and school. This means that each school can compare its students' results with the provincial, regional, and national averages. In other words, INVALSI can be an indicator of the level of a single school.

Eduscopio ranking is another evaluation system that takes into consideration the marks obtained by students in their first year of university. Each school is inserted in a rank with other schools of the same type within a radius of 30 Km centred on that school. This means that each school has a personalised rank that is useful for comparing its level with others.

The analysis of school homogeneity based on INVALSI test results and the Eduscopio ranking revealed a non-notable variability among institutions involved in our work. The percent variance in Mathematics scores for all academic tracks was around 10%, while Italian language scores show a slightly higher variability of about 12%. These values indicated that students' performance in INVALSI tests varies significantly across schools, with technical institutes showing a slightly higher level of disparity. Nevertheless, all the results were well inside the range of variability, both for the Italian Northern regions considered alone and Italy as a whole, showing a good homogeneity. A key finding is that Italian language scores were less homogeneous than Mathematics, implying differences in literacy development across schools that may not be so important for the scientific goal of this analysis.

On the other hand, the Eduscopio ranking showed that all the schools involved in our work were classified in the upper half of their ranking, suggesting that schools tended to be ranked within a narrower range when considering the results of their students during their first year at university. This means that, while standardised test scores displayed slightly greater variability, the overall educational quality, measured by the Eduscopio ranking, showed less dispersion.

Another indicator of homogeneity is the fact that all the schools in the samples had a medium-high to high socio-economic context for their students, as reported on their websites. In conclusion, considering all this information, the schools under consideration appear to share a broadly homogeneous situation.

2.4. Analysis of Effectiveness

The main idea of our research was to measure the effectiveness of our activities by checking if there was an improvement in the TOL final score before and after the attendance of the course. As explained in Section 2.2, we selected group G1, composed of 25 students who attended our course and group G2 of 99 students who did not attend. We made all the calculations described in this paper using the software R 4.5.3. Concerning the statistical test performed, we considered $p < 0.05$ as a significance level.

We started considering the average scores of both groups in their first and last attempt. As stated before, for the G1, the first TOL trial was before the attendance of the course, while the last one was after the course. It could be interesting to add that all students in the G1 group had their last attempt between May and July. This means that from the end of the course and the attempt we took into consideration, no more than 3 months have passed. As reported in Section 3, the average scores in the first and last attempts suggested that some improvement occurred in G1, but we needed to go deeper with our analysis to obtain reliable results.

First, we performed a Shapiro–Wilk test on the distribution of scores in the first and last attempts of both groups to check the normality. The results of these tests allowed us to decide which kind of statistical tools were more appropriate. We wanted to use the scores in the first attempt to evaluate and compare the beginning level. To do that, we performed a Levene's test comparing the average score obtained by the two groups in the first attempt of TOL. This was necessary to ensure the homogeneity of variance in the two groups. After that, we ran a robust independent *t*-test looking for differences between the scores in the first attempt in the two groups.

Consequently, we estimated the improvement between the first and the last attempt by performing a Yuen's trimmed mean *t*-test (Yuen, 1974) on the scores of the two trials for both groups. We decided to use the AKP as an effect size for the result of this test. We also checked the improvement, considering only the physics part of the TOL, using the same methodology just explained. In addition, we made the same calculation considering separately male and female students to check if any of our results seem to be affected by the gender of participants.

3. Results

We show here the results of the calculations described in the previous section. First, in Table 3 are reported average scores with their Standard deviation concerning the first and last attempt for G1 and G2.

Table 3. Mean scores in TOL on the first and last attempt and relative standard deviation. The minimum possible score is −25, and the maximum is 100.

Group	N	TOL 1st Attempt	STD_1st	TOL Last Attempt	STD_Last
G1	25	43.86	11.37	54.03	17.72
G2	99	50.81	12.33	60.32	12.49

These results are also represented in Figures 1 and 2. The table and the graphs suggest that the improvement in G1 is more important, but to confirm this impression, we had to perform a deeper analysis.

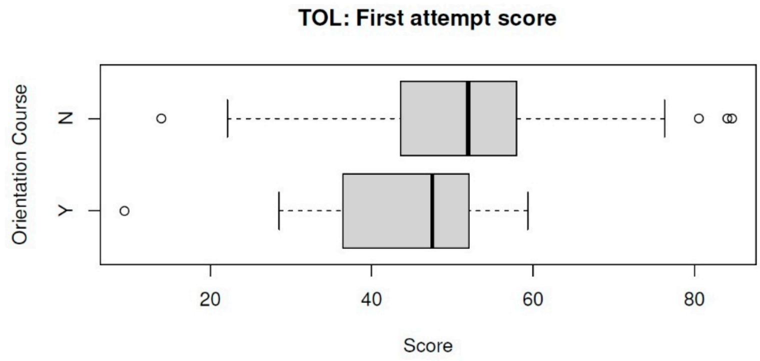

Figure 1. Standard Box-and-Whisker plot of first attempt scores for G1 (Y) and G2 (N) group.

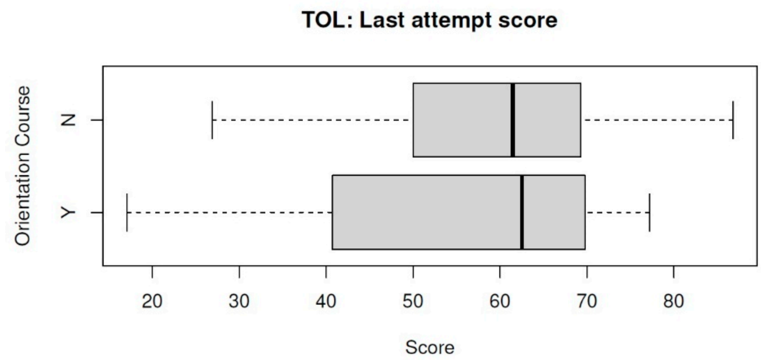

Figure 2. Standard Box-and-Whisker plot of last attempt scores for G1 (Y) and G2 (N).

As explained in Section 2.3, we checked the normality of distributions by performing a Shapiro–Wilk test on the scores of first and last attempts for both groups. We found for the G1 a p-value of 0.033 for the first and 0.058 for the last attempt. This means that the first attempt's scores distribution in G1 was not normal, while the distribution of scores in the last attempt could be considered normal, but cautiously. Concerning the G2, we obtained a p-value of 0.19 and 0.24 for the first and last attempt, respectively. This means that the distributions for the G2 group were normal. The Levene's test applied to the scores of the first attempt of G1 and G2 gave a (Pr > F) equal to 0.90 (mean centred test) and to 0.80 (median centred test). Both are much greater than 0.05, so we could confirm the homogeneity of variance.

Having normal and not-normal distributions to deal with, we decided to use robust methods to continue our analysis in accordance with (Mair & Wilcox, 2020). The robust independent t-test on the scores of the first attempt of the two groups gave a p-value of 0.047 with an effect size (AKP) of 0.54, which can be considered as "large" (Kirk, 1996). This result allowed us to claim that the initial level of G1 students was actually statistically worse than that of G2 students.

Concerning the comparison between the first and last attempts, we found a p-value of 0.0013 from Yuen's trimmed mean t-test for the G1, with an AKP effect size of 0.97 (large). This confirms that the score in the last attempt is significantly higher than in the first one for students in G1. The same calculation for the G2 gave a p-value of almost zero with an AKP of 0.82, which can be considered large. All these results are summarised in Table 4.

Considering only the physics part, we report now in Table 5 the mean scores. It is possible to notice that there probably was a significant improvement for G2, but, as explained before, we went deeper to find more reliable results

Table 4. Results of analysis concerning the scores in the TOL of G1 and G2.

Group	N	Shapiro–Wilk 1st (p-Value)	Shapiro–Wilk Last (p-Value)	Robust t-Test 1st (p-Value)	Robust t-Test 1st (AKP)	Yuen's t-Test (p-Value)	Yuen's t-Test (AKP)
G1	25	0.033	0.058	0.047	−0.54	0.0013	0.97
G2	99	0.19	0.24			~0	0.82

Table 5. Mean scores in the physics part of the TOL in the first and last attempts and relative standard deviation. The minimum possible score is −2.5, and the maximum is 10.

Group	N	Physics 1st Attempt	STD_1st	Physics Last Attempt	STD_Last
G1	25	1.66	1.25	1.84	1.65
G2	99	1.77	1.75	2.16	1.27

The Shapiro–Wilk test revealed that the distribution of scores is normal for the G1 but non-normal for the G2. We decided then to use robust methods again for both groups. In this case, we only looked for differences in the improvements, so we again performed Yuen's trimmed mean t-test on the scores of the first and last attempts for the two groups. As reported in Table 6, we found no significant improvement for students in G1 and a significant improvement for G2 with a small effect size.

Table 6. Results of analysis concerning the scores in the physics part of the TOL for G1 and G2.

Group	N	Shapiro–Wilk 1st (p-Value)	Shapiro–Wilk Last (p-Value)	Yuen's t-Test (p-Value)	Yuen's t-Test (AKP)
G1	25	0.26	0.21	0.85	null
G2	99	6.3×10^{-6}	0.038	0.016	0.24

Finally, focusing on male and female performances, we found the following results reported in Table 7.

Table 7. Results of analysis concerning the scores in the TOL of G1 and G2, considering gender.

Group	N	Shapiro–Wilk 1st (p-Value)	Shapiro–Wilk Last (p-Value)	Yuen's t-Test (p-Value)	Yuen's t-Test (AKP)
G1 M	18	0.066	0.15	0.0062	0.72
G2 M	58	0.28	0.89	2.0×10^{-5}	0.76
G1 F	7	0.46	0.069	0.023	1.07
G2 F	41	0.52	0.037	9.0×10^{-5}	0.94

Even in this case, we used robust methods because we had some distributions that were normal and others that were non-normal. Considering the p-values of the Yuen's t-test, we can see that in all groups there was a significant improvement. Concerning the female students, the effect size of G1 was 13.8% higher than that of G2, while for males, the effect sizes of G1 and G2 were very similar. In addition, the effect size for female students in G1 was 48.6% higher than that for male students in the same group.

To conclude, we did the same analysis considering only the physics part of the TOL and the results are summarised in Table 8. Observing the Yuen's p-value column, we can state that there was no significant improvement for all groups except for male students in the G2. Nevertheless, even in this case, the effect size was small, so we can assume that, concerning the physics part, there were no significant changes among the samples.

Table 8. Results of analysis concerning the scores in the physics part of the TOL for G1 and G2, considering gender.

Group	N	Shapiro–Wilk 1st (p-Value)	Shapiro–Wilk Last (p-Value)	Yuen's t-Test (p-Value)	Yuen's t-Test (AKP)
G1 M	18	0.45	0.58	0.59	null
G2 M	58	1.5×10^{-5}	0.080	0.018	0.28
G1 F	7	0.087	0.042	0.29	null
G2 F	41	0.029	0.31	0.30	0.16

4. Discussion

The results presented in Section 3 allow us to claim that both groups had a significant improvement between first and last attempts. This could be due to multiple factors. For instance, trying the TOL more than once allowed students to be more confident. In addition, after the first trial, students could be more aware of their level, and they might have prepared themselves better for further TOL sessions. More specifically, considering the effect size, we can observe that the AKP of G1 was 18.6% higher than that of G2. This suggests that students in G1 had a better improvement (RQ1). This effect could be caused by several factors. For example, it is possible that students who decided to attend our course had a higher motivation to improve their score, or maybe that, starting from a lower score in their first attempt, a significant improvement was easier for them.

Considering what we reported in this paper, we can claim that the main difference between the two groups was indeed the participation in our orientation course. As presented in Sections 2.1 and 2.2, we built the two groups carefully to ensure proper homogeneity from different points of view. In addition, the short time passed between the end of the course and the last attempt suggests that students probably did not have the opportunity to attend other courses in preparation for the test. They were certainly unable to attend other PNRR courses, as the regulations governing these activities prohibit participation in more than one edition.

It is also interesting to notice that we did not find a significant improvement concerning the physics part. Indeed, we were aware that a small number of hours was not enough to deliver a proper amount of content effectively. In addition, the physics part of the TOL is composed of just five questions, and probably these are not enough to investigate an improvement in physics knowledge and comprehension. Moreover, competencies fostered by open-ended labs may not transfer directly to the skills assessed in traditional tests. Nevertheless, the aim of our course was to foster students' skills in self-preparation. As presented at the end of Section 2, previous studies showed that OE activities help in strengthening science self-efficacy and motivation, and this leads to an improvement in general scholastic results. This is why we believe that, despite the absence of a specific improvement in physics, students were able to better prepare themselves for the TOL after attending the course (RQ2).

This last sentence seems to be particularly true for female students. Indeed, the analysis of effect sizes showed that, in the G1, female students had a much greater improvement in TOL score compared to their male counterparts. Moreover, female students in the G1 seemed to perform better than the females in the G2. We think that these differences could be a consequence of attending our course. This means that the activities we proposed seem to be more effective for girls (RQ3). Of course, because of all the limitations of this study, described in previous sections, this result cannot be interpreted as a claim of causality but just as an observational trend. Moreover, we could state that, probably, the differences in improvement described at the beginning of this section could be due mainly to the female results. Even considering the gender, a significant improvement in the physics part of the TOL was absent.

As stated in Section 2.2, we noticed that 10 students decided not to try the TOL after attending the course. We believe that even this could be considered as a positive outcome because it might mean that the course helped those students realise that an engineering programme was not the right choice for them.

All these results allow us to claim that the use of OE laboratories in an orientation course can have positive outcomes. Students attending this type of course may develop valuable skills that enable them to prepare independently and effectively for the university entrance exam. Moreover, such activities can support students in making informed decisions regarding their future university pathways, which may contribute to lower dropout rates in higher education.

This course has also been offered in subsequent academic years. Therefore, the analysis of data from later editions may confirm or refine the conclusions drawn from the data presented in this paper. More precisely, considering several editions in different years, we could work with a larger sample, obtaining more significant results. In addition, thanks to some improvements to the material used during the course (in particular in the questionnaire used in the introductory meeting during subsequent editions), we could obtain more information about students' level and comprehension.

Author Contributions: Conceptualization, M.Z. and M.B.; methodology, M.B. and R.V.; software, M.B., R.V., M.C.; validation, M.B.; formal analysis, M.B., R.V.; investigation, All authors; resources, M.Z.; data curation, All authors; writing—original draft preparation, R.V.; writing—review and editing, R.V.; visualization, R.V.; supervision, M.Z.; project administration, M.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and with the Code of Ethics and Conduct of Politecnico di Milano (https://www.polimi.it/fileadmin/user_upload/Il-Politecnico/statute-regulations/Codice_Etico_e_di_comportamento_del_Politecnico_di_Milano_-_modifica-it_it-en_gb-IR-C.pdf, accessed on 10 November 2025).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The complete dataset is available on the Zenodo website at the following URL: https://doi.org/10.5281/zenodo.17472842.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

OE	Open-Ended
PBL	Problem-Based Learning
FCI	Force Concept Inventory
TOL	Test On-Line
PNRR	National Recovery and Resilience Plan

References

Alhadabi, A. (2021). Science interest, utility, self-efficacy, identity, and science achievement among high school students: An application of SEM tree. *Frontiers in Psychology*, *12*, 634120. [CrossRef] [PubMed]

Bandura, A. (1997). *Self-efficacy: The exercise of control* (pp. ix, 604). W H Freeman/Times Books/Henry Holt & Co.

Bodner, G. M. (1986). Constructivism: A theory of knowledge. *Journal of Chemical Education*, *63*(10), 873. [CrossRef]

Bozkurt, A. (2020). Educational technology research patterns in the realm of the digital knowledge age. *Journal of Interactive Media in Education*, *2020*(1), 18. [CrossRef]

Bozzi, M., Mazzola, R., & Zani, M. (2023). Peer learning in higher education: An example of practices. *Il Nuovo Cimento C*, *46*(6), 206. [CrossRef]

Bozzi, M., Raffaghelli, J. E., & Zani, M. (2021). Peer learning as a key component of an integrated teaching method: Overcoming the complexities of physics teaching in large size classes. *Education Sciences*, *11*(2), 67. [CrossRef]

Celik, P., Onder, F., & Silay, I. (2011). The effects of problem-based learning on the students' success in physics course. *Procedia-Social and Behavioral Sciences*, *28*, 656–660. [CrossRef]

Craig, N. C., & Hill, C. S. (2012). Do-it-yourself experiments for the instructional laboratory. *Journal of Chemical Education*, *89*(6), 755–758. [CrossRef]

De Backer, L., Van Keer, H., & Valcke, M. (2015). Exploring evolutions in reciprocal peer tutoring groups' socially shared metacognitive regulation and identifying its metacognitive correlates. *Learning and Instruction*, *38*, 63–78. [CrossRef]

DM 934/2022. (2022). Available online: https://www.mur.gov.it/it/atti-e-normativa/decreto-ministeriale-n-934-del-03-08-2022 (accessed on 1 March 2025).

Doğru, M. S., & Kurnaz, M. A. (2023). Students' contextualizing knowledge on the mirage incident, reflection and refraction: A case study. *Physics Education*, *58*(6), 065004. [CrossRef]

Doucette, D., Clark, R., & Singh, C. (2018, December 31). *What's happening in traditional and inquiry-based introductory labs? An integrative analysis at a large research university*. 2018 Physics Education Research Conference Proceedings, Washington, DC, USA. Available online: https://www.per-central.org/items/detail.cfm?ID=14775 (accessed on 10 November 2025).

Edström, K., & Kolmos, A. (2014). PBL and CDIO: Complementary models for engineering education development. *European Journal of Engineering Education*, *39*(5), 539–555. [CrossRef]

Gondoni, P., Bozzi, M., Balossi, B., Calisesi, G., Ferocino, E., Genco, I., Molteni, L. M., Pini, A., Rezoagli, F., Zanoletti, M., & Zani, M. (2021, March 8–9). *A novel approach to online physics refresher courses at politecnico di milano*. 15th International Technology, Education and Development Conference, Online. [CrossRef]

Gök, T. (2011). Exploring of students' performances, motivation processes and learning strategies in studio physics. *Latin American Journal of Physics Education*, *5*, 154–161.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, *66*(1), 64–74. [CrossRef]

Hattie, J. (2023). *Visible learning: The sequel: A synthesis of over 2100 meta-analyses relating to achievement*. Routledge. Available online: https://www.routledge.com/Visible-Learning-The-Sequel-A-Synthesis-of-Over-2100-Meta-Analyses-Relating/Hattie/p/book/9781032462035 (accessed on 10 November 2025).

Hattie, J., & Clarke, S. (2018). *Visible learning: Feedback*. Routledge. [CrossRef]

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, *30*(3), 141–158. [CrossRef]

Hoellwarth, C., Moelter, M. J., & Knight, R. D. (2005). A direct comparison of conceptual learning and problem solving ability in traditional and studio style classrooms. *American Journal of Physics*, *73*(5), 459–462. [CrossRef]

Holmes, N. G., & Wieman, C. E. (2018). Introductory physics labs: We can do better. *Physics Today*, *71*(1), 38–45. [CrossRef]

How to access to Laurea programmes in engineering. (2025). Available online: https://www.polimi.it/en/prospective-students/how-to-apply/admission-to-laurea-programmes/engineering (accessed on 10 January 2025).

Istenič, A. (2024). Blended learning in higher education: The integrated and distributed model and a thematic analysis. *Discover Education*, *3*(1), 165. [CrossRef]

Jones, L. (2007). *The student-centered classroom*. Cambridge University Press.

Jungert, T., Hubbard, K., Dedic, H., & Rosenfield, S. (2019). Systemizing and the gender gap: Examining academic achievement and perseverance in STEM. *European Journal of Psychology of Education*, *34*(2), 479–500. [CrossRef]

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, *56*(5), 746–759. [CrossRef]

Kirkwood, A., & Price, L. (2014). Technology-enhanced learning and teaching in higher education: What is 'enhanced' and how do we know? A critical literature review. *Learning, Media and Technology*, *39*(1), 6–36. [CrossRef]

Lam, P. L. C., Ng, H. K. Y., Tse, A. H. H., Lu, M., & Wong, B. Y. W. (2021). eLearning technology and the advancement of practical constructivist pedagogies: Illustrations from classroom observations. *Education and Information Technologies*, *26*(1), 89–101. [CrossRef]

Mair, P., & Wilcox, R. (2020). Robust statistical methods in R using the WRS2 package. *Behavior Research Methods*, *52*(2), 464–488. [CrossRef]

Marsh, H. W., Van Zanden, B., Parker, P. D., Guo, J., Conigrave, J., & Seaton, M. (2019). Young women face disadvantage to enrollment in university STEM coursework regardless of prior achievement and attitudes. *American Educational Research Journal*, *56*(5), 1629–1680. [CrossRef]

Marušić, M., & Sliško, J. (2014). Students' experiences in learning physics: Active learning methods and traditional teaching. *Latin-American Journal of Physics Education*, *8*, 32.

Mazzola, R., Gondoni, P., Bozzi, M., Scagliola, A., Pagnoni, S., & Zani, M. (2021, November 8–9). *Physics experiments and peer-assessment: An high-school—University project*. 14th Annual International Conference of Education, Research and Innovation, Online. [CrossRef]

Otto, S., Bertel, L. B., Lyngdorf, N. E. R., Markman, A. O., Andersen, T., & Ryberg, T. (2024). Emerging digital practices supporting student-centered learning environments in higher education: A review of literature and lessons learned from the COVID-19 pandemic. *Education and Information Technologies*, *29*(2), 1673–1696. [CrossRef]

Pintrich, P. R., & Others, A. (1991). *A manual for the use of the motivated strategies for learning questionnaire (MSLQ)*. Available online: https://eric.ed.gov/?id=ED338122 (accessed on 10 November 2025).

Raffaghelli, J., Ghislandi, P., Sancassani, S., Canal, L., Micciolo, R., Balossi, B., Bozzi, M., Di Sieno, L., Genco, I., Gondoni, P., Pini, A., & Zani, M. (2018). Integrating MOOCs in physics preliminary undergraduate education: Beyond large size lectures. *Educational Media International*, *55*, 301–316. [CrossRef]

Redish, E. F., Saul, J. M., & Steinberg, R. N. (1997). On the effectiveness of active-engagement microcomputer-based laboratories. *American Journal of Physics*, *65*(1), 45–54. [CrossRef]

Smith, E. M., & Holmes, N. G. (2021). Best practice for instructional labs. *Nature Physics*, *17*(6), 662–663. [CrossRef]

Trumper, R. (2003). The physics laboratory—A historical overview and future perspectives. *Science & Education*, *12*(7), 645–670. [CrossRef]

Vickrey, T., Rosploch, K., Rahmanian, R., Pilarz, M., & Stains, M. (2015). Research-based implementation of peer instruction: A literature review. *CBE—Life Sciences Education*, *14*(1), es3. [CrossRef]

Virzi, R., Bozzi, M., Costigliolo, M., & Zani, M. (2025a, June 17). *Comparison between Rasch analysis and Classical Test Theory for a Physics questionnaire validation*. Eleventh International Conference on Higher Education Advances, Valencia, Spain. [CrossRef]

Virzi, R., Bozzi, M., Mazzola, R., Costigliolo, M., & Zani, M. (2025b). *ST2 TOL/orientation course 2022/2023 dataset* [Data set]. Zenodo. [CrossRef]

Virzi, R., Bozzi, M., Mazzola, R., & Zani, M. (2025c). Orientation activities for high school students: Propaedeutic test and open-ended physics lab. In I. Testa, M. Michelini, & S. Esposito (Eds.), *Connecting physics education research and practice* (pp. 205–215). Springer Nature Switzerland. [CrossRef]

Wilcox, B. R., & Lewandowski, H. J. (2016). Open-ended versus guided laboratory activities: Impact on students' beliefs about experimental physics. *Physical Review Physics Education Research*, *12*(2), 020132. [CrossRef]

Wood, A. K. (Ed.). (2023). *Effective teaching in large STEM classes*. IOP Publishing. [CrossRef]

Yuen, K. K. (1974). The two-sample trimmed t for unequal population variances. *Biometrika*, *61*(1), 165–170. [CrossRef]