This document is the Accepted Manuscript version of a Published Work that appeared in final form in Connecting Physics Education Research and Practice book, copyright ©2025 Springer Nature Switzerland AG after peer review and technical editing by the publisher. To access the final edited and published work see https://dx.doi.org/10.1007/978-3-031-86609-8_13.

Orientation activities for high school students: propaedeutic test and open-ended Physics lab

R. Virzi[0009−0001−3465−9300], M. Bozzi[0000−0002−8117−6379],
R. Mazzola[0000−0002−9188−0503], and M. Zani[0000−0001−7960−3362]

Abstract As part of the project of Italian Ministry of University and Research related to active orientation from school to university, among the courses proposed by the Politecnico di Milano was the one titled "Methodological Introduction to Engineering Test Preparation". The aim of this course was to provide methodological tools for studying Physics and Mathematics topics and developing logical-mathematical skills, scientific methods, modeling and problem-solving. The course was structured into modules covering motivational aspects, Mathematics and Physics. The latter consisted of an introductory online lesson on learning aspects, followed by an in-person laboratory activity. During the introductory session, students completed a multiple-choice questionnaire, aiming to highlight gaps or misconceptions that typically make the Engineering Test challenging. The laboratory activity was designed following an open-ended lab model. Students were encouraged to explore, with maximum autonomy, the physical quantities related to a certain phenomenon and discover the relationships between them. The research activity reported here focuses on validating the preparatory questionnaire

1 Introduction

During the academic year 2022/2023, the Politecnico di Milano offered a wide range of orientation courses for high school students. These courses were designed in accordance with the guidelines outlined in the National Recovery and Resilience Plan (PNRR)[1], ensuring specific criteria in terms of duration and content [2]. To be more specific, these courses were structured to span 15 hours, aiming primarily to assist students in exploring and comprehending the university environment while fostering awareness of their expectations and aspirations for future studies. Among these offerings was our orientation course titled *Introduzione metodologica alla*

Roberto Virzi e-mail: roberto.virzi@polimi.it · Matteo Bozzi · Roberto Mazzola · Maurizio Zani
Department of Physics, Politecnico di Milano, Milan, Italy

preparazione al test di Ingegneria (Methodological Introduction to Engineering Test Preparation). As implied by its name, the course targeted high school students seeking to equip themselves for the Engineering Test named *Test On Line* (TOL) [3]required for admission to the Engineering faculties at Politecnico di Milano University.

The course was structured in different moments. 3 hours were dedicated to an in-presence lesson focused on motivation. The remaining 12 hours were equally divided into more specific activities of Physics and Mathematics. Our group provided the Physics part organizing 2 hours of an online introductory lesson during the which students were asked to answer to an 8-items propaedeutical physics test. After that, students were asked to participate in a 4-hour laboratory session. In total, 113 students have joined the proposal. The whole course was conducted in Italian.

At the end of this experience, we had the possibility to analyze all the data of participants applying the Classical Test Theory[4] in order to validate our questionnaire and the effectiveness of distractors. In the following sections are presented all the results of these analyses and also some consideration on how we could improve the test for future editions according to these results.

All these results are presented in the following sections to give an answer to the research question: is our test reliable and useful for the purposes of our orientation course?

2 The open-ended style laboratory

This work is particularly focused on the analysis of the propaedeutical test but before to go deeper in this topic it's useful to know something more about the structure of the laboratory. The main aim of our laboratory activity was to help students in developing skills in order to be able to discover and study physical contents in an autonomous way. These aims could be achieved, for instance, by asking students to discuss in their group how to solve an issue they found during their work or how to choose which quantities are useful and what should be a proper method to obtain a reliable measure.Considering this we preferred to avoid a classical approach and choose an open-ended style [5]. This means that students did not receive detailed instructions about how to perform an experiment but only a very general description of the phenomenon they were supposed to investigate and some suggestions concerning which physical quantities they should consider. Each group (normally no more than 4 people per group) had to work on a different topic and had access to very simple material to set up the experiment. During the laboratory session, students could rely on the support of our research group which was composed of an associated professor, a research fellow, a PhD candidate, and a graduating student. At the end of these four hours, each group was asked to present their work to the rest of the class. This different approach has a positive influence because it allows students to feel more like protagonists of their work and makes them aware of the experimental nature of Physics [6]. Moreover, it enhances the development of critical thinking and fosters the science self-efficacy.Self-efficacy is related to how a person believes in his capability to perform a task[7]. In particular, science self-efficacy is very important for the learning process because it is strictly related to science interest and science achievement[8] that are fundamental for an orientation activity. It has also been observed that self-efficacy can help students to cope with more challenging tasks with less stress[9]. Recent studies also highlighted that self-efficacy can affect the results of students in Physics exams[10]. In addition, we think that this methodology can stimulate student's curiosity and help them to develop an interest for science knowledge and science learning that could be important in their future path [11].

To give a proper idea of how our activities were organized we provide here a brief explanation of a laboratory session. First, students were divided into small groups composed by 3, 4, or a maximum of 5 persons per group. Then, each group received a box with different materials and a card containing very general information about the phenomena they were asked to study. An example of a card is shown in figure 1

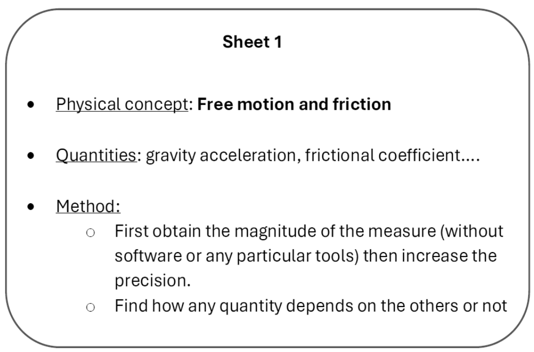


Fig. 1 Example of a card containing initial information to perform the laboratory experience.

We prepared 7 cards covering the following topics: motion and friction, oscillations, pendulum, motion in a viscous medium, density and volume, optics, mechanical energy conservation. Then the teacher gave a brief introduction and after that students could start their work. During the lab, teachers continuously stimulated students to find solutions and invent strategies: these dialogues were useful for giving and receiving feedback. In the end, each group of students presented the results of their work, giving reason for their choices and answering possible questions from teachers or coursemates

3 The propaedeutical questionnaire

The propaedeutical questionnaire that students faced in the introductory online lesson was useful to help them become aware of their gaps in the knowledge of Physics

contents. The 8 questions forming our test were designed carefully [12] and share conceptual similarities, for example, with those found in the CSEM test [13] or in the FCI[14]. Each item follows a structure with 4 possible answers, wherein only one is correct, while the other 3 are intentionally designed to identify potential gaps or misconceptions. The topics covered by the 8 items are summarized in Table 1.
In this questionnaire, each participant received 1 point for each correct answer and 0 point for each wrong or missing answer.

Table 1 List of topics covered by each item

Item	Topic
Q1	Force and movement
Q2	Friction
Q3	Circular motion
Q4	Kinetic and potential energy
Q5	Electric field and electric potential
Q6	Period and angular velocity
Q7	Free fall
Q8	Electrostatic force and Newton's third law

Furthermore, after completing the test, the instructor who conducted the introductory lesson provided the correct answers, accompanied by comments on incorrect choices, and sought feedback from participants. This aspect was important in prompting students to reflect on their knowledge and learning methods, fostering a meta-cognitive approach[15].

According to Hattie and Clarke (2019)[16] feedback from teachers is fundamental in learning process especially when help students providing a direction to follow for their improvement. In the opposite direction, feedback from students can make the learning visible by teachers. In our situation, the feedback given by the instructor was focused on the misconceptions highlighted by the test in order to help students to develop skills useful for recognizing and avoiding them in the future. As for the feedback from students, given that the course was at the beginning, it was not related to what the students learned, but it provided useful information about the reasons for choosing a particular distractor. In this case, feedback could help students become aware of their mistakes and conscious of the way they learn and memorize.

Considering that the introductory meeting was online, this questionnaire was submitted using the Socrative [17] platform.

4 Test validation

The validation is a fundamental step to be sure that our test is actually useful for our purposes[18] or if it needs some changes. In our evaluation, we referred to several different coefficients that can provide useful information about single questions and the whole test as summarized by L. Ding and R. Beichner in 2009[4]. We could rely on data related to 113 students coming from 11 different schools and overall from 13 different school types.

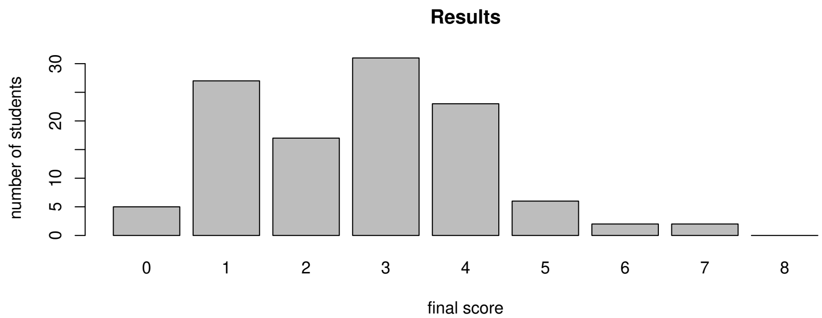


Fig. 2 Total score of the whole group of students.

Above all, we can observe in Fig 2 that only very few students reached a score greater than the 50% of the total while the most part of them collected between 1 and 4 points up to 8. The average score of the sample is 2.67. It is particularly low but we want to underline that the aim of this test was not to evaluate the knowledge of a single student but to obtain an overview of gaps and misconceptions of the whole group. We can also notice that no one answered correctly to all questions. In this section we'll show the values of each coefficient we considered in our analysis.

Table 2 Coefficient related to each item: Difficulty Index (P), Discrimination Index (D) and Point Biserial Coefficient. (r_{pbi})

Item	P	P-Level[19]	D	D-Level [12]	r_{pbi}
Q1	0.32	Medium	0.48	Good	0.45
Q2	0.38	Medium	0.56	Good	0.45
Q3	0.19	Medium-High	0.47	Good	0.51
Q4	0.28	Medium	0.58	Good	0.52
Q5	0.25	Medium-High	0.33	Reasonably good	0.30
Q6	0.50	Medium-Low	0.50	Good	0.37
Q7	0.58	medium-Low	0.55	Good	0.44
Q8	0.20	Medium-High	0.27	Marginal	0.27

In Table 2 are reported the Difficulty Index (P), the Discrimination Index (D), and the Point Biserial Coefficient (r_{pbi}) that provide an indicator for each item.

The difficulty index (P) is defined as the ratio between the number of students that answered correctly to a question and the total amount of students:

$$P_n = \frac{N_1}{N}$$

this means that a high difficulty level states that a high number of students answered correctly. In other words, the higher the difficulty index, the easier the question.

The value of P is useful even to calculate the point biserial coefficient defined as:

$$r_{pbi} = \frac{\overline{X_1} - \overline{X_0}}{\sigma_x}\sqrt{P(1-P)}$$

where $\overline{X_1}$ is the average test score of students that answered correctly to a particular question, $\overline{X_0}$ is the average test score of students that answered wrong to the same question, σ_x is the standard deviation of scores and P is the difficulty level of the question. The point biserial coefficient expresses the reliability of a single question, in other words, it indicates the correlation between the item score and the total score. All our results can be labeled as *acceptable* because they are all above 0.2 [20].
The Discrimination Index provides information about how much an item is able to recognize students with an overall high performance. To calculate it we used the following formula:

$$D_n = \frac{N_H - N_L}{N/4}$$

in which N_H is the number of students on the upper quartile that answered correctly to the question Q_n, N_L is the number of students in the lower quartile that answered correctly to the same question and N is the total number of students. Considering that our sample was composed of 113 persons and that the possible score ranged from 0 to 8 we found many situations of different students with the same score. This made hard to choose which student were to consider in the upper quartile and which in the lower quartile.
To fix this problem we decided, using the R language[21] to write a script to sort randomly 1000 times the students with the same score and calculate the Discrimination Index for each possibility. The values reported in Tab 2 are the average of these 1000 values of D.

Concerning P and D is reported in Table 2 also the level related to each value [19, 12]. As far as r_{pbi} concerns, all the values can be labeled as *acceptable* because they are all above 0.2 [20]. Observing the table we can notice that 3 questions had a medium-high difficulty ($P \leq 0.25$) while 2 had a medium-low difficulty ($0.75 \leq P < 0.50$). Concerning D it's clear that all the items are good or reasonably-good ($D \geq 0.30$) while only the Q8 was a marginal item ($0.20 \leq D < 0.30$). It's interesting to observe that Q3 has been a difficult question nevertheless it had a good discrimination index.

The last two coefficient we took in consideration were the Kuder-Richardson reliability (r_{test}) and the Ferguson's delta (δ) as shown in Tab 3.

The Kuder-Richardson reliability is calculated with

$$r_{test} = \frac{K}{K-1}(1 - \frac{\Sigma P_i(1-P_i)}{{\sigma_x}^2})$$

Table 3 Coefficient related to the whole test: Kuder-Richardson reliability (r_{test}) and the Ferguson's delta (δ)

Coefficient	Value
r_{test}	0.32
δ	0.90

where K is the number of item in the test, P_i is the difficulty level of question i and σ_x is the standard deviation of total scores. r_{test} expresses the internal consistency of the whole test.
Eventually, the Ferguson's delta can be obtained with

$$\delta = \frac{N^2 - \Sigma {f_i}^2}{N^2 - N^2/(K+1)}$$

where N is the number of students, K is the number of items and f_i is the frequency of students with a total score i. δ indicates an overall discrimination level.

The value of r_{test} is 0.32 and it indicate that the test has a low reliability and it could be used for survey and group averages but not to assess a single student[22]. Nevertheless, as we wrote before, the aim of our test was exactly to underline gaps and misconceptions of the whole group of students so that a low Kuder-Richardson reliability was expected. in addition, r_{test} measures the internal coherence of a test, and remembering Table 1, it's clear that in a test in which each question covers a different topic it's hard to have a high internal coherence. On the contrary, the value of δ is high so our test provides a good level of discrimination among students [20].

Eventually, we focused on the effectiveness of the distractors in each question. The Table 4 shows all the percentages of choice for the four possible answers for each item. The percentage corresponding to the right answer is written in bold.

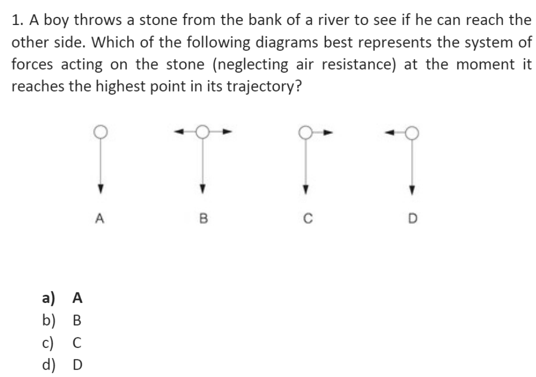
Table 4 Percentages related to the choice of each answer. The percentage corresponding to the right answer is written in bold.

Answer/Item	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
A	**31.5**	42.0	62.8	19.6	33.6	21.2	5.4	66.1
B	9.9	2.7	**18.6**	12.5	20.9	**50.4**	33.9	8.0
C	49.5	**38.4**	2.7	**27.7**	**25.5**	8.8	**58.0**	6.3
D	9.0	17.0	15.9	40.2	20.0	19.5	2.7	**19.6**

Observing the table it is possible to notice that only 3 distractors were chosen by less than 5% of the students so the most part of distractors can be considered effective. More specifically, in 5 items there is a distractor that was chosen more often than the actual right answer, highlighting a very good effectiveness and suggesting the possible presence of a misconception.
According to these results, we made a general review of our test to be more effective for future editions of the course. In the next section, we show our changes

5 Questionnaire review

Taking into account the results obtained in Section 4 and the feedback received by students during the online lesson, we decided to make some modifications to Items 1, 2, 3, and 7.
Figure 3 shows the text of the first question. Remembering Tables 2 and 4 we can


1. A boy throws a stone from the bank of a river to see if he can reach the other side. Which of the following diagrams best represents the system of forces acting on the stone (neglecting air resistance) at the moment it reaches the highest point in its trajectory?

a) A
b) B
c) C
d) D

Fig. 3 Question 1 of our questionnaire. In bold the right answer.

state that Q1 is generally a good item. Nevertheless, we observed that answers C and D are equivalent because the text does not specify if the stone is thrown from left to right or vice-versa. The most chosen answer was C and we think that the reason is that students probably interpreted the missing text as *from left to right* and applied the common misconception *motion implies active force*[14] or *impetus dissipation*[14]. To avoid this possible misunderstanding we specified the thrown direction (left to right) and we changed the answers as shown in Figure 4

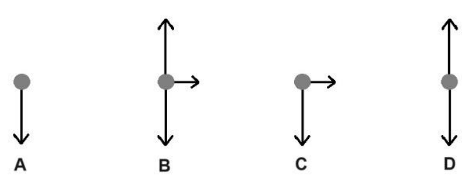


Fig. 4 New set of possible answers to item 1

We expect that students who choose the answers B or D think that at the maximum point, to have zero vertical velocity, a zero net force is required.

Figure 5 shows the second item. In that case, we wanted to improve the answer B because it was a poor distractor chosen by less than 5% of students. The new

2. Applying a horizontal force of F_1=5 N to an object placed on a horizontal surface, it is observed that the object does not move due to friction. Even when a horizontal force of F_2=10 N is applied, the object still does not move because of friction. What is the magnitude of the frictional force acting on the object?

a) It is greater than 10 N
b) It is equal to 10 N
c) **It is 5 N in the case of F_1 and 10 N in the case of F_2**
d) It cannot be determined with the given data

Fig. 5 Question 2 of our questionnaire. In bold the right answer.

answer B is “it does not depend on the applied force” so we also had to adjust the last sentence of the item changing it to “which statement related to friction force acting on the object is correct?”. We think that this new distractor could detect a common misconception about friction: *Friction is constant and equals to $f = \mu N$ (normal force)* [23].

3. A body moves in uniform circular motion along a circumference with a radius of 1 m at a speed of 5 m/s. Which statement is correct?

a) Since it is uniform motion, the speed is constant and the acceleration is zero.
b) The tangential acceleration is zero, and the centripetal acceleration is 25 m/s^2
c) The tangential acceleration is zero, and the centripetal acceleration is 4 m/s^2
d) None of the above is correct.

Fig. 6 Question 3 of our questionnaire. In bold the right answer.

Even in the third item, shown in Figure 6, we had a poor distractor because answer C was chosen only by 2.7% of students. We decided then to change it to “The centripetal acceleration is approximately 4m/s^2”. This new formulation aims to detect students who are aware that there is an acceleration greater than zero but who calculate it wrongly by dividing speed by the period.

7. If you drop an object without giving it any push, it falls with an acceleration of 9.8 m/s^2, neglecting air resistance. However, if you throw it downward with force, what will its acceleration be after it leaves your hand, still assuming there is no air resistance?

a) It is less than 9.8 m/s^2
b) It is greater than 9.8 m/s^2
c) It is equal to 9.8 m/s^2
d) It depends on the initial velocity of the object

Fig. 7 Question 7 of our questionnaire. In bold the right answer.

Eventually, we focused on item 7 to improve the effectiveness of distractors A and D . Starting from the item shown in Figure 7, we changed answer A to “It varies during the fall” and answer D to “It depends on the force with which the object is thrown". These two new answers involve a misconception about gravity: *Constant gravity does not imply constant acceleration*[23] and the already cited *impetus dissipation*[14]

6 Conclusion

As seen in Section 4, the Classical Test Theory confirms that our questionnaire was well-structured and reliable. Concerning Table 4 we can observe that in the first 5 questions and in the last one students chose a distractor more often than the right answer and this could be a hint of the presence of specific misconceptions related to the topics covered by those questions. In addition, we can notice that only 3 "wrong answers" were chosen by fewer than 5% of the students so that most of our distractors were effective. This allowed us to repeat this experience during the academic year 2023/2024 with the corrections exposed in Section 5.
In addition, in 2023/2024 we transformed the questionnaire into a 2-way test adding a Likert-scale question after each item. in this new question, we asked "how much do you feel confident about your answer?" with 4 possible choices. We think that this will help us in better recognize misconceptions of students.
We also started to estimate the effectiveness of our course by observing the results of our students in the actual TOL and the first results seem to be interesting. In addition, we are analyzing the results in the updated propaedeutic questionnaire performing a new validation

References

1. Decreto ministeriale 934/2022 (2022)
2. Corsi di orientamento attivo nella transizione scuola università (2023). URL https://www.polimi.it/orientamento-pnrr
3. What the test consists of and how to prepare (2024). URL https://www.polimi.it/en/futuri-studenti/admission-to-laurea-programmes/engineering/what-the-test-consists-of-and-how-to-prepare
4. L. Ding, R. Beichner, Physical Review Special Topics-Physics Education Research **5**(2), 020103 (2009)
5. R. Trumper, Science & Education **12**, 645 (2003)
6. B.R. Wilcox, H. Lewandowski, Physical Review Physics Education Research **12**(2), 020132 (2016)
7. X. Hu, Y. Jiang, H. Bi, International Journal of STEM Education **9**(1), 47 (2022)
8. A. Alhadabi, Frontiers in Psychology **12**, 634120 (2021)
9. A. Bandura, *Self-efficacy: The exercise of control.* (New York: W.H. Freeman and Company, 1997)
10. A. Malespina, C. Singh, European Journal of Physics (2024)
11. Z. Hazari, G. Potvin, R.H. Tai, J. Almarode, Physical review special topics-physics education research **6**(1), 010107 (2010)
12. R.L. Ebel, D.A. Frisbie, *Essentials of educational measurement* (Prentice-Hall Englewood Cliffs, NJ, 1972)

13. D.P. Maloney, T.L. O'Kuma, C.J. Hieggelke, A. Van Heuvelen, American Journal of Physics **69**(S1), S12 (2001)
14. D. Hestenes, M. Wells, G. Swackhamer, et al., The physics teacher **30**(3), 141 (1992)
15. M. Romainville, Studies in Higher Education **19**(3), 359 (1994)
16. J. Hattie, S. Clarke, *Visible Learning: Feedback* (Routledge, 2019)
17. Socrative. URL https://www.socrative.com/
18. D.L. Bandalos, *Measurement theory and applications for the social sciences* (Guilford Publications, 2018)
19. L. Crocker, J. Algina, *Introduction to classical and modern test theory.* (ERIC, 1986)
20. P. Kline, *A Handbook of Test Construction: Introduction to Psychometric Design* (Methuen, 1986)
21. The R project for statistical computing. URL https://www.r-project.org/
22. R.L. Doran, *Basic Measurement and Evaluation of Science Instruction.* (ERIC, 1980)
23. G. Liu, N. Fang, International Journal of Engineering Education **32**(1), 19 (2016)