

Article

# Peer Learning as a Key Component of an Integrated Teaching Method: Overcoming the Complexities of Physics Teaching in Large Size Classes

Matteo Bozzi <sup>1,\*</sup> , Juliana E. Raffaghelli <sup>2</sup> and Maurizio Zani <sup>1</sup> <sup>1</sup> Department of Physics, Politecnico di Milano, 20133 Milan, Italy; maurizio.zani@polimi.it<sup>2</sup> Department of Psychology and Educational Sciences Studies, Universitat Oberta de Catalunya, 08018 Barcelona, Spain; jraffaghelli@uoc.edu

\* Correspondence: matteo.bozzi@polimi.it

**Abstract:** Over the last decade, policy makers have urged universities to innovate their teaching methodologies. Although educational research has shown that active methods lead to improvements in learners' performance more than traditional lectures in small classes, some factors impede active methods from spreading in large size classes. In this paper we aim at fostering these methodological innovations by showing the effectiveness of an integrated teaching methodology that employs peer learning, technology, and traditional lectures in large size classes. In the academic years 2017–2018 and 2018–2019 a quasi-experiment involved more than 600 engineering students per year attending an introductory physics course at Politecnico di Milano. These learners were assigned to two sections and their outcomes in a pre-test and a post-test were analyzed through descriptive and inferential statistics. The learning results of the experimental group were always better than the outcomes of the control group, regardless of the difficulty of the physics topics addressed. Furthermore, a very low threshold for the exposure to peer learning has been highlighted in order for it to be effective. These promising results may thus foster ongoing changes in university policy towards the renewal of the teaching methodology.

**Keywords:** peer learning; physics teaching; higher education; large size classes; active learning; traditional lectures



**Citation:** Bozzi, M.; Raffaghelli, J.E.; Zani, M. Peer Learning as a Key Component of an Integrated Teaching Method: Overcoming the Complexities of Physics Teaching in Large Size Classes. *Educ. Sci.* **2021**, *11*, 67. <https://doi.org/10.3390/educsci11020067>

Academic Editor: David Arendale  
Received: 29 December 2020  
Accepted: 4 February 2021  
Published: 9 February 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

### 1.1. Introduction

Physics educational research has highlighted that learning physics appears to be a noteworthy and challenging testing ground for a large number of university learners enrolled in science, technology, engineering and mathematics (STEM) faculties [1–3]. By way of illustration, 2350 engineering students enrolled in a physics course held at the Politecnico di Milano during the first term of the academic year 2018–2019, but only less than six students out of ten passed their own final examination before the beginning of the second academic term. Their average mark was lower than 24/30, the highest mark being 30/30 cum laude in Italy [4].

The majority of the physics courses are characterized by a pedagogy centered on traditional lectures [5]. However, traditional lectures have been highly criticized for being a teacher-centered strategy that focuses on teaching itself rather than on learning. An evocative metaphor for illustrating the role played by the instructors in this transmittal model is a “sage on the stage” [6], whereby educators transfer knowledge to their students, who act as passive learners and play the role of empty boxes to be filled. A classic lecture, however, does not seem to perfectly combine with diversified learning modes and backgrounds [7–9]. This could be the alleged cause of failure to achieve satisfactory learning and results in academic physics courses for many students. To reinforce this idea, it should be

emphasized that quality teaching, which is crucial to assure a quality higher education [10], should be characterized by the opposite priority, i.e., it should pursue learning rather than teaching [11].

In contrast to the aforementioned traditional teacher-centered pedagogy, since the early 1990s new active learning strategies have been a focus of discussion [12,13]. According to Vygotsky's theory [14], every learner is characterized by a "zone of proximal development", a distinctive unreal area where learning does not take place individually but occurs through other people's assistance. Interactions among human beings facilitate and mediate higher mental functioning within zones of proximal development. Coherently to the educational approach suggested by constructivism, learners have to build their own knowledge and be the protagonists of their learning path. Many studies demonstrate that students' learning improves if they are actively engaged during their lessons [13,15–17]. Furthermore, student-centered pedagogy increases class attendance, students' acquisition of expert attitudes towards the discipline and student engagement [18–20]. In a student-centered pedagogy the role of the instructor undergoes a deep-seated change: from a "sage on the stage" the instructor becomes a "guide on the side" [6]. Peer Instruction [15], Problem-Based Learning [12] and Student-Centered Activities for Large Enrollment Undergraduate Programs [21] methods have been developed in some North American universities in the context of basic physics courses. In these educational methodologies students are put at the center of the learning and teaching process and every student is engaged. The engagement of all learners is precisely the essential element that allows to classify an educational strategy as active learning [16].

However, it seems that there are complicating factors in the adoption of active methods that prevent their spread and trust amongst university teachers. In fact, these active learning strategies are actually not characterized by large-scale use in the STEM field, even though they are globally spread, albeit unevenly. Although the employment in physics courses has truly taken hold in a not negligible, though still restricted, number of US academic institutions [5], this is not the case of European universities, let alone Italian ones [22–24].

Firstly, there are a number of common reasons frequently indicated by instructors to explain their resistance to active learning strategies, such as lack of time, limited resources and a dearth of university or departmental support, along with some problems concerning syllabus content coverage [25,26]. Secondly, learners exposed to a student-centered educational strategy show undeniable improvements in conceptual understanding, but these improvements do not always correspond to as many gains in their problem-solving skills [27,28]. As a result, active methods require further investigation to confirm their specific impact and effectiveness. Thirdly, one of the most challenging issues in undertaking active methods relates to their complex logistics in large size lectures. Increasingly large classes, resulting from the massification of higher education [29–31], face the trade-off of simpler logistics through traditional lectures at the cost of lower acquisition of knowledge and skills, owing to the further reduced chance for the students to interact with their instructors and receive feedback from them [32].

Overcrowded classrooms have been considered a hindrance in several research studies where large size classes (LSC) appear to be frequent factors allegedly making university instructors adopt traditional lectures as the only feasible pedagogical approach [5,33,34]. Incidentally, a large size class may correspond to a different number of students depending on the disciplines and pedagogical needs of the learning environment being considered. For instance, in the Fine Arts, fifteen students may represent a large size class, whereas a first-year physics class may be defined as large if there are 100 learners or more ( $N \geq 100$ ). Regardless of any numeric threshold, a large class appears to be an environment where the quality of learning may be negatively impacted by the number of learners. Moreover, research on active learning carried out in Europe frequently involves small ( $N < 50$ ) or medium ( $50 \leq N < 100$ ) size classes [28,35,36], so university lecturers have little evidence regarding how to devise active pedagogies in large class formats. As a part of the active

strategies, peer learning (PL) is probably deemed disadvantageous in large size classrooms, in spite of its value in supporting effective learning in higher education [37].

To summarize, the sudden transition of academic physics courses from a completely teacher-centered pedagogy to a totally student-centered one, particularly in the context of LSC, appears to be hardly feasible and not entirely satisfactory in the short term. It is thereby appropriate and urgent to explore the educational effectiveness of the synergistic and integrated use of traditional lectures and PL—a specific active learning model—augmented by the use of technology. Moreover, it would be interesting to investigate effective ways of implementing such an approach in LSC. In this regard, in an Italian scenario dominated by academic physics courses centered on transmittal lectures, Politecnico di Milano has begun to work on active learning in large class formats, with promising results contributing to a broader picture that requires further investigation [38].

## 1.2. Background

### 1.2.1. Active Methods, Peer Learning and Large Size Lectures in Higher Education

Human beings tend to build their own mental models in order to organize their experiences and observations [9,39]. Not only are these models instruments to interact with reality, but they also determine how individuals assimilate new information and experiences [7]. According to Ausubel's studies [40–42], it is appropriate to identify and distinguish two antithetical human learning modes: rote learning and meaningful learning. There is a continuum between these two learning modes owing to the different degrees of development of people's relevant cognitive structure as well as their dissimilar attempts to include new conceptual meanings [43].

However, mental structures and processes are deeply connected with social and cultural ones on account of the interaction among people and between individuals and collective culture: learning is indeed a social process [14,44]. As a consequence, since learning in groups ought to be more efficacious than learning as isolated individuals [45], student-centered strategies like PL appear to be particularly suitable in this regard. Aligning with the cognitivist and the socio-cultural theory of learning, empirical research in higher education has differentiated between deep learning and surface learning [46]. If the former is characterized by a critical analysis of new ideas and facts, connecting them with existing cognitive structures and building innumerable links among these ideas, the latter incorporates them uncritically, tending to store them as isolated, detached items [47,48].

All in all, the theory and research about learning in higher education are pointing out that active methods are an effective approach to trigger meaningful, situated and deep learning [23,48]. In the context of STEM faculties, active methods have been originally developed in the US, where notable efforts have been made to change undergraduate education [49,50]. Afterwards, the attention to the STEM disciplines and the modernization of educational strategies have progressively spread also in Europe [51], Italy included [52–54].

Although a unique definition has not been synthesized, scholars do agree on the fact that an active method is any instructional methodology that engages students in the learning process and requires them to perform meaningful learning activities during classroom time, while encouraging learners to reflect on what they are doing [55]. Nowadays several lines of research suggest that not only are these active strategies able to improve learners' performance more than passive ones [13,15–17], but they make a substantial contribution to overcoming the gap in learning between under-represented minority students and non-under-represented minority learners, thus triggering the growth in science self-efficacy for all learners [56]. Along with the growing evidence that these interactive-engagement methodologies are effective, there are increasing calls to train more and more instructors in the use of these educational strategies in their academic courses [57].

Nevertheless, the effective use of active strategies requires an accurate design of the learners' activities and teamwork; inadequate logistics and work organization risk undeniably compromising the students' learning [58]. These just mentioned criticalities are

more noteworthy in the context of academic large classes, where interactive-engagement methodologies are sporadically employed [59] and seem to be rarely successful [60].

In the context of active methods, PL appears to be particularly interesting. In a salient article, Hattie [59] synthesized more than 1200 meta-analyses that cover about 65,000 researches and over 80 million students ranging from early childhood to the tertiary level. He classified 195 factors that are related to learning outcomes from very positive influences to very negative effects, and their average effect size (Cohen's  $d$ ) was 0.40. Hattie's study [59] highlighted that some popular active methodologies show a positive effect size, but lower or equal to the mean value; for instance, problem-based learning Cohen's  $d$  and cooperative learning Cohen's  $d$  were, respectively, 0.12 and 0.40. On the contrary, factors like classroom discussion ( $d = 0.82$ ), positive peer influences ( $d = 0.53$ ) and peer tutoring ( $d = 0.53$ ) evidenced an effect size that is higher than the average value. Similarly, analyzing the findings of 98 single studies, the meta-analyses of both Bowman-Perrott et al. [61] and Leung [62] demonstrated significant effects of peer tutoring on different students age groups.

Contrary to popular belief, PL does not refer to a single, undifferentiated educational methodology: over the years ten different typologies have been recognized [63]. However, beyond the different models, PL is considered a form of interdependent learning characterized by the sharing of knowledge, ideas and experience among participants, thus being mutually beneficial [64]. A further crucial feature is that all peers play the same role: nobody has a role as teacher or expert practitioner, even though some peers could have appreciable experience and expertise; consequently, no one has power over others owing to their status [65]. Actually, PL is characterized by a horizontal and non-hierarchical relational status.

PL settings are likely to produce fertile instructional dialogues among peers as they support joint problem solving, depend on intrinsic rather than extrinsic rewards and discourage competition among students [66]. A line of research has focused on investigations of the possible improvements in the students' outcomes when they are exposed to PL or, more generally, to active learning in a student-centered learning environment rather than in a traditional classroom. Notwithstanding that some studies highlight better students' results in active learning spaces rather than in traditional classrooms [67,68], some other studies lead to the conclusion that their findings are not dissimilar in the two learning settings [69,70]. However, classroom renovation can be extremely expensive and new technology may rapidly become obsolete [68,70]. In most cases, however, the learning environment is traditional and assigned by the institution, and thus no change is possible. As a result, not only are active methods generally employed in assigned structural contexts, but their implementation is further contrasted by the frequent occurrence of LSC.

As a matter of fact, class size plays a pivotal role owing to its influence on students' attainments, learning and teaching quality, and program evaluation. University stakeholders give considerable attention to this issue on account of its economic implications [71]. Actually, as a consequence of the economic crisis in 2008, government funding allocated for higher education has progressively decreased both in the US [72] and in the European Union [73]: an increase in class size reduces human resources as well as operating costs. In addition, massification of higher education [29–31] has propelled academic institutions to that solution.

However, a traditional teacher-centered pedagogy is normally ineffective or blandly successful when applied to LSC [74], while the adoption of student-centered strategies and the use of technology succeed in improving learners' outcomes [33,75–77]. With regard to the use of active methods, research aiming at investigating the influence of class size on student achievement generally indicates better results in small groups [78,79].

### 1.2.2. Peer Learning in Physics

PL is employed in some different courses of action with reference to the teaching of physics [35,80]. For instance, in peer instruction, an active learning technique introduced

by Harvard professor E. Mazur in 1991 [15], an initial short presentation on an essential topic supplied by the instructor is followed by a conceptual question on that item with different possible answers. The query aims at investigating students' understanding of that theme and every student responds individually to that question. The students compare and discuss their answers with their neighbors for a few minutes, explaining their reasoning, before being given a chance to answer singly a second time. Finally, the instructor shows and discusses the correct answer to the question and moves on to the next topic.

Differently from this approach, the active methodology used by the physics Education Research Group at the University of Massachusetts, called class-wide discussion [81], begins with a small group peer discussion of a conceptual question, followed by an individual or group answer. Afterwards, the undergraduates participate in a class-wide discussion fostered by the instructor, who favors discussion across peer groups by asking different students to share their reasoning with the class. Furthermore, the Physics Education Group at the University of Washington has developed a set of tutorials that are meant to boost physics teaching at an introductory level [82]. Taking place after the issue has been illustrated in lectures, tutorial sessions consist in small peer group discussions guided by a specific tutorial worksheet. The students find their own answers to the allotted tasks through interaction with their peers as well as the tutorial instructors.

The overall exposure time of the learners to active instructional practices in physics teaching is impressively heterogeneous between studies, and thus the total class time devoted to one of these student-centered strategies may range from 10% to 100% (class completely taught with active learning) [17]. Furthermore, this exposure time is substantially considered as a fixed parameter rather than a variable to investigate, i.e., there are few inquiries aimed at evaluating whether and how it influences the educational strategy success. In other words, what is the minimum time devoted to PL during a course in order for appreciable learning to be achieved by the students?

Conversely, many studies have compared the effectiveness of interactive-engagement methodologies and traditional lectures in relation to physics teaching. Some wide and significant meta-analyses [13,17] emphasize that active learning is more effective, particularly with regard to conceptual understanding [27,28]. Nevertheless, this students' learning assessment is frequently performed through the same standardized concept inventories [83], like the Force Concept Inventory [84], and thus the effectiveness of the educational strategy as a function of the assessment test difficulty, i.e., the complexity of the concepts checked, is not investigated.

In this regard, a significant but neglected area of research focuses on investigating PL effectiveness with specific reference to the complexity of both the physics issues addressed by students and the tasks assigned to them. This is an even more tricky challenge if the size of the class is large and the efficacy of active learning in academic physics teaching appears to be substantially different whether small or large size classes are engaged. For instance, Freeman et al. [17] consider Hedges's  $g$  as an effect size and emphasize that its mean value is 0.635 (medium) and 0.314 (small) in the cases of respectively small or LSC involved in the analyzed study.

However, a recent study focused on physics refresher courses allotted to future freshmen before the beginning of the academic year has pointed out that the synergistic and integrated use of massive open online courses, PL and traditional lectures allows the students to achieve positive learning outcomes regardless of the different size of the classes; in fact, no differences in the effect size have been highlighted [38].

Since research focused on LSC may not be receiving proper attention, particularly with regard to the European context [28], further investigations appear to be needed and appropriate on the integration of PL into traditional lectures, as a possible solution to the problems posed by LSC in academic physics courses.

## 2. Materials and Methods

### 2.1. Research Design

In light of the above considerations, the authors identified the most appropriate investigation methodology. The students involved could not be assigned randomly to the experimental group or the control one owing to some ethics and organizational constraints of our undergraduate course setting. As a consequence, a quasi-experiment was considered as an appropriate methodological approach [85,86], which tends to be a frequent option in complex educational settings [87,88].

Designed to be implemented in an ongoing physics course activity, i.e., in a natural environment, in the academic years 2017–2018 and 2018–2019, the quasi-experiment aimed at investigating the effectiveness of the integration of PL activities, strengthened by the use of technology, into traditional physics lectures as a teaching method in LSC.

The research questions were:

- (RQ.1) Is an integrated teaching methodology (ITM) based on the mixed use of PL, technology and lecture, an effective strategy in order to learn physics in university LSC?
- (RQ.2) Is this ITM more effective than the usual transmission form consisting of traditional academic lectures?
- (RQ.3) Is there a threshold for the exposure to PL (as a key component of ITM) in LSC in order for it to be effective? If so, what is this minimum level?
- (RQ.4) Is this ITM successful despite the level of difficulty of the physics topics being tested?

### 2.2. Research Context and Participants

The research activity was implemented in an academic course on mechanics and electromagnetism called “Fisica Sperimentale A + B” (“Experimental Physics A + B”) at Politecnico di Milano, located in the North-Western part of Italy. In a traditional approach, “Fisica Sperimentale A + B” lasted for 100 h, 60 h of which consisted of traditional lectures focused on a theoretical framework, while 40 h were devoted to drills in order to develop and heighten the students’ problem-solving skills. This physics program is usually provided to 19-year-old freshmen attending the first year of materials and nanotechnology engineering and those studying chemical engineering during the first term of each academic year.

The students enrolled were 674 (153 females, 380 males and 141 unidentified) and 610 (159 females, 398 males and 53 unidentified), respectively, in the academic years 2017–2018 and 2018–2019. Owing to these large numbers, the students were divided into three different sections on the basis of an alphabetical order. These cohorts were assigned to the same three instructors in both academic years, maintaining the teaching approaches constant. For the purposes of this study, one of the groups, with the same teacher, was identified as the studied group (SG), whereas the other two sections (with two different but constant instructors) represented the overall control group (CG).

The attendance of the course was not mandatory, based on the academic rules, and the students were not compelled to attend it, coherently with the choice of carrying out our research in a natural environment. However, although some freshmen did not attend the lessons regularly, the average number of students participating in each group was about 100 or higher; as a consequence, the size of the classes could be classified as large.

### 2.3. Learning Design and Assessment Tools

The aforementioned traditional design of the course was confirmed for the CG, while it was modified for the SG, which experienced periodical PL sessions. Data on SG and CG students’ starting level in physics and on their possible equivalence in terms of academic performance were gathered through a multiple-choice test arranged in one of the earliest lessons. This initial test consisted of twelve quizzes focused on classical physics studied at high schools and, at the same time, pertaining to some topics taught in the academic physics courses of Politecnico di Milano; the time allotted to each quiz, characterized by

four possible alternatives of which only one was correct, was two minutes. The freshmen earned one point for each correct answer and nil in all other cases; the overall score attained by a freshman was normalized so that the lowest and the highest possible value were equal to zero and ten, respectively.

In order to evaluate the effectiveness of the educational strategies adopted in the SG and in the CG, a final multiple-choice test was administered to the students towards the end of the course (three months later). The time devoted to each quiz and the method of quantifying the performance of every freshmen in this final test were identical to the initial one. Furthermore, the structure of this assessment was equivalent to the initial test, as it consisted of twelve quizzes directly related to the topics covered in the syllabus.

Although both the initial and final tests administered to the freshmen in academic years 2017–2018 and 2018–2019 consisted of twelve questions, these tests were modified from one year to the next. In 2017–2018 quizzes were meant to evaluate the students' conceptual understanding of some significant topics that they had studied in their university physics course. However, the authors also aimed at investigating the effectiveness of their ITM with relation to the complexity of the physics issues that characterized the test. Therefore, in the second year of the quasi-experiment, the initial and final test quizzes were changed and focused on some possible misconceptions or misunderstandings in physics, related to some significant topics that students had previously explored in high school and that they would study at the university. As stated by the scientific literature on this issue, misconceptions appear to be extremely widespread and persistent [89,90]; consequently it was likely that the test could have been more difficult in 2018–2019 than in the previous year. It appears that in the second year the freshmen had to face a double challenge: to demonstrate their conceptual understanding of some meaningful topics and, in addition to this, to overcome any misconceptions regarding those subjects.

In order to verify and quantify the different characteristics between the tests administered in 2017–2018 and 2018–2019, a classical item analysis was carried out [91]. The test difficulty index (DI) was meaningfully different in the two academic years and equal to 0.41 and 0.26, respectively; however, based on Crocker and Algina [92] the difficulty of these tests can be classified as average. Conversely, their discrimination index (DS) was similar, respectively 0.35 and 0.31 in 2017–2018 and 2018–2019, and could be classified as reasonably good [91].

Unlike the CG, the SG experienced some PL sessions during their lesson time. In this new approach, "Fisica sperimentale A + B" lasted for 100 h. Of this total amount, about 57.5 h consisted of lectures, while about 2.5 h were devoted to PL sessions and 40 to drills. Figure 1 shows the learning design of the SG.

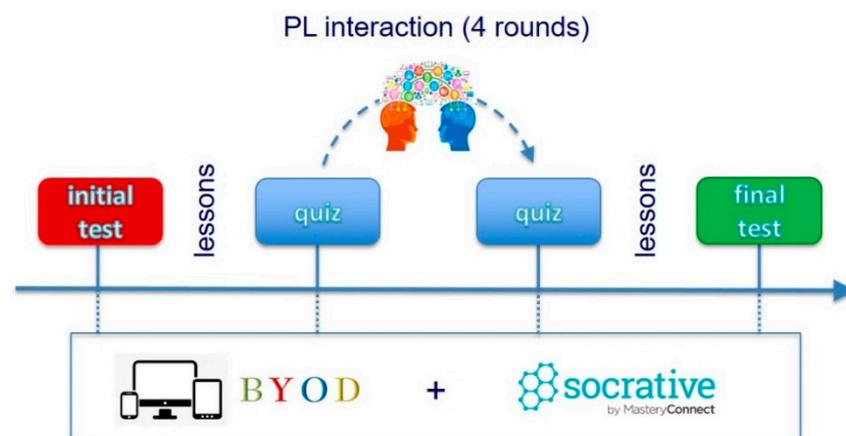


Figure 1. Learning design.

During each PL session, the students individually answered some brief multiple-choice quizzes; both the first and the second PL periods were based on mechanics topics

and consisted of five quizzes each, whereas the third and the fourth PL period focused on electromagnetism issues and the students answered four quizzes in each session. These intermediate quizzes differed from the multiple-choice quizzes of the initial and final tests, in that they focused on other topics. The quizzes selected for the PL sections were the same in the two academic years encompassed by our quasi-experiment. As a result, it should be emphasized that these intermediate quizzes had been specifically designed for the quasi-experiment implemented in the first academic year and they were not centered around possible misconceptions. Consequently, their effectiveness could be inferior in 2018–2019 and this might be a limitation of our study.

SG freshmen answered these quizzes through the student response system (SRS) Socrative, whose use has been widely discussed [93,94] in the implementation of interactive teaching. Coherently to the documented positive results of the Bring Your Own Device (BYOD) strategy [95], the SG students employed their own electronic devices, like laptops, smartphones and tablets. It has been pointed out that the use of technology within higher education enhances student engagement [96] and students' academic performance [97,98], and PL particularly benefits from its use [97]. Each PL session was characterized by the following schedule:

- Answering the first quiz individually (2 min);
- Peer discussion in small groups (3–5 min);
- Without getting any feedback from the other classmates and the teacher, SG freshmen were asked to work with their neighboring fellow students so as to discuss briefly the quizzes in small groups (three or four freshmen);
- Answering individually the same quiz again (2 min);
- . . . . . ;
- Answering the last quiz individually (2 min);
- Peer discussion in small groups (3–5 min);
- Answering individually the same quiz again (2 min);
- Finally, showing the correct answer to every question and the rate of answers ascribed to each possible alternative, the teacher gave a brief explanation (2–3 min).

Table 1 summarizes the PL session time.

**Table 1.** Peer learning session time.

PL Session	Number of Quizzes	PL Session Time (min)	Overall Time (min)	Peer Discussion in Small Groups Time (min)
1	5	37–48		15–25
2	5	37–48		15–25
3	4	30–39	134–174	12–20
4	4	30–39		12–20

Finally, each of the tests administered to the students, i.e., the initial and final test and every intermediate quiz were previously evaluated and considered adequate by a panel of instructors and experts.

#### 2.4. Data Analysis

Data gathered in this research were preliminarily explored through descriptive statistics. Afterwards, a Shapiro–Wilk test allowed to check if they were normally distributed and a Levene test was performed to test their homoscedasticity. All detailed preliminary analyses and outcomes can be found as open data for transparency [99], and therefore they have not been reported in their entirety in the following “Results” section.

Since our data did not fit the normal distribution, a further inferential analysis was carried out by means of non-parametric tests, like the Mann–Whitney U test for unpaired data, the Wilcoxon signed-rank test for paired data with one independent factor, and the Kruskal–Wallis test for more than one factor. The statistical significance was chosen at the

level  $\alpha = 0.01$ , the effect size adopted was Cohen's  $d$  and its value was classified according to Sawilowsky [100]. Finally, data analysis was carried out through the statistical open source software RStudio (<https://rstudio.com/> accessed on 26 December 2020).

### 3. Results

A preliminary investigation consisted in highlighting and comparing the level of knowledge in physics of the SG and the CG both before and after the quasi-experiment through descriptive statistics. We analyzed the performance of all the students who took part in a specific test (between subject design) as well as the results of the freshmen who participated in both initial and final tests (within subject design). Table 2 summarizes these findings for the different sections.

**Table 2.** Descriptive score statistics (between and within subject design) for the different sections in initial (Ini) and final (Fin) tests.

Section	Academic Year	Type of Data	Students Number		Median		Mean		Standard Deviation	
			Ini Test	Fin Test	Ini Test	Fin Test	Ini Test	Fin Test	Ini Test	Fin Test
SG	2017–2018	Between subject	128	99	4.17	6.67	3.57	6.68	1.46	1.44
CG	2017–2018	Between subject	303	228	4.17	5.83	4.29	5.75	1.79	1.94
SG	2017–2018	Within subject	63	63	4.17	6.67	3.68	6.87	1.39	1.45
CG	2017–2018	Within subject	133	133	4.17	5.83	4.61	5.83	1.67	1.76
SG	2018–2019	Between subject	137	103	2.50	3.33	2.87	3.48	1.43	1.63
CG	2018–2019	Between subject	322	265	3.33	2.50	3.11	2.94	1.31	1.56
SG	2018–2019	Within subject	84	84	2.50	3.33	2.83	3.55	1.30	1.70
CG	2018–2019	Within subject	180	180	3.33	3.33	3.30	3.04	1.21	1.48

Regardless of the type of data, the studied freshmen showed a lower mean normalized score than the control freshmen in the initial test along with better results in the final one in both academic years. Moreover, while the SG outcomes improved between the initial and final test in 2017–2018 as well as in 2018–2019, the CG performance in the final test was worse than in the initial one in 2018–2019.

Furthermore, the mean score achieved by both the studied and the control students in the initial test as well as in the final one in the academic year 2018–2019 was remarkably inferior to the mean score reached in the previous year. These different outcomes were not related to a change in data collecting methods, a change in personnel, or a local/global event, but to the choice of focusing the tests in the second year on some possible misconceptions.

The low scores generally achieved by all the students in the 2018–2019 initial test demonstrate that what we tested were misconceptions broadly held by freshmen, whereas the scores achieved by them in the final test confirm the persistence of these misconceptions.

#### 3.1. RQ.1 (and RQ.4): Overall ITM Effectiveness

To determine whether ITM is an effective strategy in order to learn physics in university LSC, the findings of the experimental section freshmen who took both the initial and final tests were analyzed. Both academic years 2017–2018 and 2018–2019 were considered.

Since the Shapiro–Wilk and Levene's tests did not support a normal distribution and homoscedasticity ( $p < 0.01$ ) [99], a non-parametric Wilcoxon signed-rank test for paired data was employed. Statistically significant differences between the initial and final test were found. The corresponding effect sizes were both interesting, even though very different, and the 99% confidence interval did not include zero. This meant that the improvements in physics of the SG students were not ascribable to randomness and their intensity was appreciable. Table 3 shows these findings.

These results allowed to partially answer RQ.1, in terms of evidence supporting that ITM based on the use of PL, technology and lectures can be effective when teaching physics in LSC.

**Table 3.** Wilcoxon signed-rank test results for SG in the academic years 2017–2018 and 2018–2019 (within subject design).

Section	Academic Year	Type of Data	Students Number		Wilcoxon Signed-Rank Test			
			Ini T	Fin T	V	p-Value	d	CI <sub>99</sub>
SG	17–18	Within subject	63	63	2012	$5.995 \times 10^{-12} \ll 0.01$	2.24 (huge)	1.65–2.84
SG	18–19	Within subject	84	84	1914.5	0.001877 $\ll 0.01$	0.48 (small)	0.071–0.89

### 3.2. RQ.2 (and RQ.4): ITM vs. Traditional Academic Lecture Effectiveness

Having implemented Shapiro–Wilk and Levene’s tests [99], a comparison between the SG and the CG at the beginning of the physics course was completed by using a Mann–Whitney U test. Firstly, considering all the students of both the studied and the control section who took the initial test, the *p*-value called the attention to statistically significant differences between the knowledge in physics of the two sections in 2017–2018: the CG was definitely better. On the contrary, in 2018–2019 the two groups appeared to perform in an equivalent way, with no significant statistical differences in their mean score. Secondly, taking into account only the students who took both the initial and final tests, the *p*-value highlighted that the CG seemed to reveal a significantly higher initial level of knowledge. Table 4 summarizes the test outcomes.

**Table 4.** Mann–Whitney U test results in initial test for SG and CG in academic years 2017–2018 and 2018–2019 (between and within subject design).

Section	Academic Year	Type of Data	Test	Students Number	Mann–Whitney U Test			
					Z	p-Value	d	CI <sub>99</sub>
SG	17–18	Between subject	Ini	128				
CG	17–18	Between subject	Ini	303	3.6577	$10^{-4} \ll 0.01$	0.43 (small)	0.15–0.70
SG	17–18	Within subject	Ini	63				
CG	17–18	Within subject	Ini	133	3.2636	$4 \times 10^{-4} \ll 0.01$	0.59 (medium)	0.18–0.99
SG	18–19	Between subject	Ini	137				
CG	18–19	Between subject	Ini	322	1.8339	0.0657 > 0.01 (> 0.05)		
SG	18–19	Within subject	Ini	84				
CG	18–19	Within subject	Ini	180	2.8875	0.0047 $\ll 0.01$	0.38 (small)	0.33–0.72

Thirdly, in order to check the SG and the CG learning at the end of the quasi-experiment, the results of the freshmen taking part in the final trial were compared by means of a Mann–Whitney U test for unpaired data. It showed that the SG achieved statistically better outcomes than the CG in both academic years, even though the effect size showed that the intensity of the effect was negligible in academic year 2018–2019. Table 5 summarizes the test outcomes.

**Table 5.** Mann–Whitney U test results in final test for SG and CG in academic year 2017–2018 and 2018–2019 (between subject design).

Section	Academic Year	Type of Data	Test	Students Number	Mann–Whitney U Test			
					Z	p-Value	d	CI <sub>99</sub>
SG	17–18	Between subject	Fin	99				
CG	17–18	Between subject	Fin	228	−4.0804	$10^{-4} \ll 0.01$	0.97 (large)	0.72–1.21
SG	18–19	Between subject	Fin	103				
CG	18–19	Between subject	Fin	265	−3.0112	0.0028 $\ll 0.01$	−0.040 (negligible)	−0.25–0.17

Finally, a Wilcoxon signed-rank test for paired data was carried out to examine the results of the freshmen who completed both the initial and final tests. Comparing the Wilcoxon signed-rank test implemented for the SG, shown previously in Section 3.1, and the same test with specific reference to the CG, one can emphasize that the SG always

showed statistically significant differences between initial and final knowledge and the effect size was always higher than the effect size that characterized the CG. Indeed, the control section knowledge appeared to be better in the final test than in the initial one only in academic year 2017–2018; on the other hand, no statistical differences were highlighted between the two test outcomes with reference to 2018–2019. Table 6 shows the test findings.

**Table 6.** Wilcoxon signed-rank test results in initial and final test for SG and CG in academic years 2017–2018 and 2018–2019 (within subject design).

Section	Academic Year	Type of Data	Students Number		Wilcoxon Signed-Rank Test			
			Ini T	Fin T	V	p-Value	d	CI <sub>99</sub>
SG	17–18	Within subject	63	63	2012	$6.0 \times 10^{-12} \ll 0.01$	2.24 (huge)	1.65–2.84
CG	17–18	Within subject	133	133	5502	$2.9 \times 10^{-10} \ll 0.01$	0.71 (medium)	0.38–1.04
SG	18–19	Within subject	84	84	1914.5	$0.001877 \ll 0.01$	0.48 (small)	0.071–0.89
CG	18–19	Within subject	180	180	4554.5	$0.06563 > 0.01$		

These outcomes allowed to answer RQ.2, in terms of evidence supporting that ITM is more effective than the usual transmission form with relation to physics learning in LSC.

### 3.3. RQ.3 (and RQ.4): Threshold for Exposure to PL for Effective Learning in LSC

Despite belonging to the SG, some of the studied freshmen did not attend all of the PL sessions; in fact, some of them were absent when the PL sessions took place. Table 7 shows their actual participation in these lectures.

**Table 7.** Number of students who attended the PL sessions in the academic years 2017–2018 and 2018–2019.

Section	Academic Year	PL Sessions				
		4	3	2	1	0
SG	2017–2018	41	20	10	6	22
CG	2017–2018	0	0	0	0	228
SG	2018–2019	43	29	17	3	11
CG	2018–2019	0	0	0	0	265

After performing Shapiro–Wilk and Levene’s tests [99], a Kruskal–Wallis test was conducted to take into consideration the different numbers of PL sessions that the students had taken part in. This test aimed at finding a threshold for the exposure to PL in LSC in order for it to be effective. The group characterized by no attendance at PL sessions included freshmen belonging to both the SG and the CG in order to reduce the effect of the instructors on the freshmen’s findings and increase the number of members in the benchmark group.

Considering the academic year 2017–2018, the Kruskal–Wallis test revealed a significant effect of the number of PL sessions on the students’ outcomes (chi-squared = 19.423, df = 4,  $p = 0.0006491 \ll 0.01$ ). Notwithstanding that this test showed that a difference existed, it was impossible to establish where it laid; as a consequence, manual post hoc Mann–Whitney U pairwise comparisons were carried out and statistically significant differences were found between four and zero PL sessions (4–0:  $p = 0.00648 \ll 0.01$ ; 4–1:  $p = 1.000$ ; 4–2:  $p = 0.576$ ; 4–3:  $p = 1.000$ ; 0–3:  $p = 0.0308 > 0.01$  and  $< 0.05$ ; 0–2:  $p = 1.000$ ; 0–1:  $p = 1.000$ ; 1–2:  $p = 1.000$ ; 1–3:  $p = 1.000$ ; 2–3:  $p = 0.458$ ). However, it should be highlighted that the test between zero and three PL sessions showed a very small  $p$ -value, which was statistically significant the level  $\alpha = 0.05$ .

Moreover, in order to further decrease the possible instructor effect on the undergraduates’ learning, the SG students who took both the initial and final tests were arranged

on the basis of the attended PL sessions. A Wilcoxon signed rank test was implemented and the effect size was calculated. Table 8 summarizes these data and results with specific reference to 2017–2018.

**Table 8.** SG students who took both the initial and final tests in 2017–2018, arranged on the basis of the PL sessions attended: descriptive score statistics in final test and Wilcoxon signed rank test.

PL Sessions	Students Number	Median		Mean		Standard Deviation		V	Wilcoxon Signed Rank Test	
		Ini Test	Fin Test	Ini Test	Fin Test	Ini test	Fin test		p	d
4	36	4.17	6.67	3.81	7.04	1.37	1.31	666	$1.667 \times 10^{-7} \ll 0.01$	$2.40 > 2.24 = d_{\text{overall}}$
3	14	4.17	7.50	4.05	7.02	1.03	1.41	104	$0.00120 \ll 0.01$	$2.41 > 2.24 = d_{\text{overall}}$
2	9	4.17	6.67	2.96	6.02	1.91	1.71	45	$0.00763 \ll 0.01$	$1.68 < 2.24 = d_{\text{overall}}$
1	4	-	-	-	-	-	-	-	-	-
0	0	-	-	-	-	-	-	-	-	-

The inferential test appeared to be statistically significant for all levels of PL that were possible to check. It is worth highlighting that the effect size changed evidently between two and three PL sessions; furthermore, in comparison with the same test carried out on the overall studied section, characterized by  $d_{\text{overall}} = 2.24$  and illustrated in Section 3.1, it should be underlined that the effect size corresponding to two PL sessions was considerably lower than that “average”  $d_{\text{overall}}$  (75%). On the other hand, in the case of three or four PL sessions the effect size was slightly higher (7%).

An identical analysis was implemented with reference to the academic year 2018–2019. The Kruskal–Wallis test did not show a significant effect of PL sessions attendance on the students’ outcomes at level  $\alpha = 0.01$  (chi-squared = 11.896,  $df = 4$ ,  $p = 0.0182 > 0.01$  and  $< 0.05$ ). However, the  $p$ -value was very small and it would be considered statistically significant at the level  $\alpha = 0.05$ . Coherently, the manual post-hoc Mann–Whitney U pairwise comparisons did not highlight statistical significance at the level  $\alpha = 0.01$ .

Finally, a Wilcoxon signed rank test was performed to investigate the results of the SG freshmen who took both the initial and final tests. Table 9 synthesizes these data and results with specific reference to 2018–2019.

**Table 9.** SG students who took both the initial and final tests in 2018–2019, arranged on the basis of the PL sessions attended: descriptive score statistics in final test and Wilcoxon signed rank test.

PL Sessions	Students Number	Median		Mean		Standard Deviation		V	Wilcoxon Signed Rank Test	
		Ini Test	Fin Test	Ini Test	Fin Test	Ini Test	Fin Test		p	d
4	40	2.50	3.33	2.85	3.60	1.36	1.76	424.5	$0.0293 < 0.05$	$0.48 = 0.48 = d_{\text{overall}}$
3	28	3.33	3.75	2.83	3.27	1.27	1.67	226	$0.1978 \gg 0.05$	-
2	12	2.50	3.75	2.78	3.81	1.44	1.61	39.5	$0.2207 \gg 0.05$	-
1	2	-	-	-	-	-	-	-	-	-
0	2	-	-	-	-	-	-	-	-	-

The inferential test did not appear to be statistically significant for all levels of PL that were possible to check at  $\alpha = 0.01$ . However, the  $p$ -value corresponding to four PL sessions was small and statistically significant at the level  $\alpha = 0.05$ . With specific reference to this case, it is worth highlighting that the effect size was equal to the effect size that characterized the overall studied section ( $CI_{95}$ : from 0.025 to 0.93) and illustrated above in Section 3.1.

These results allow to answer RQ.3, in terms of evidence supporting that there is a threshold for the exposure to PL in LSC in order for it to be effective, which corresponds to three or four sessions at the level  $\alpha = 0.05$ .

### 3.4. RQ.4: Effectiveness of ITM and Topic Difficulty

RQ.4 could be answered by aggregating the results reported in RQ.1, RQ.2 and RQ.3. The answer to RQ.1 highlighted the ITM's effectiveness with reference to physics learning in academic LSC. As a matter of fact, it was verified that this educational method was effective both in 2017–2018 and 2018–2019, notwithstanding that the size of the effect was different in the two academic years.

As regards RQ.2 it was found that the ITM was more effective than the usual transmission form in learning physics in LSC. In fact, this claim appears to be appropriate with reference to both academic years, even though the size of the effect was higher in 2017–2018.

Finally, with reference to RQ.3 it was emphasized that there is a threshold for the exposure to PL in LSC in order for it to be effective. Needless to say, this result characterized both academic years even though this threshold is higher when the level of difficulty of the physics issues addressed is greater.

In conclusion, this ITM tended to be successful despite the level of difficulty of the physics topics being tested.

## 4. Discussion

A quasi-experiment was conducted to test the effectiveness of a new educational practice as a physics teaching method in university LCS throughout the academic years 2017–2018 and 2018–2019. Comparing the results of the SG and the CG students in the final test, statistically significant differences in their physics conceptual understanding in favor of the experimental section were highlighted, regardless of the level of difficulty of the physics topics tested.

Owing to the low number of PL sessions experienced by the SG students, it was a priori unlikely that a better understanding of a few physical concepts, covered in the PL sessions, might determine an overall better knowledge of all topics studied in their physics course. Rather, the PL sessions aimed at creating a positive learning experience fostering students' self-esteem, self-efficacy and the development of self-assessment. These factors are considered crucial in order for the students to achieve overall meaningful learning. Moreover, this possible positive learning experience seemingly promoted a different study method for participants outside the formal context of the lectures, i.e., students might employ a PL strategy in their own study time.

Although the use of active methods in the context of academic LSC is a major challenge and their efficacy in learning physics appears to be moderate on average, the intensity of the effect measured was definitely higher than the mean values mentioned in the scientific literature (Hedges'  $g = 0.314$  [17], Cohen's  $d = 0.40$  [58]). When searching for moderators of this effect, it was not possible to uncover any influence from the instructor or the level of difficulty on the undergraduates learning at this stage. Therefore, a plausible conclusion was that the ITM is effective in relation to learning physics in university LSC (RQ.1), performing better than traditional methods (RQ2).

One might consider at this point the specificities of the ITM's effect; the RQ3 and RQ4 come to help in this endeavor. In fact, the overall evidence showed that ITM is effective with a relatively small time of exposure to PL (RQ3) and regardless of the level of difficulty of the issues tested (RQ.4).

With reference to the third research question (RQ.3), data analysis seems to highlight the existence of a threshold for the exposure to PL, as a key component of ITM, in large class formats in order for it to be effective. From inferential analysis this minimum level seems to be higher when the trial difficulty is greater and this threshold could be established at about 110 min (three PL sessions) and 150 min (four PL sessions) in 2017–2018 and 2018–2019, respectively.

Focusing on the studied students who were grouped on the basis of the attended PL sessions and who took both the initial and final tests, not only could this study strengthen the aforementioned conclusion, but it also succeeded in neutralizing the contribution

of the teacher, who was always the same, to the freshmen's learning. Not every group showed statistically significant differences between the initial and final tests; moreover, the corresponding intensities of the effect were not all equal to each other and to the "average" value characterizing the overall SG. Consequently, ITM, with a strong presence of PL mediated by technologies like SRSs, seems to be effective regardless of the instructor's contribution to the undergraduates' learning.

Moreover, if only the students who participated both in the initial and final tests are considered, the aforementioned claim on the ITM's greater efficacy relative to the traditional teaching strategy is further corroborated. More evidence comes from the case of all of the students who took the final trial: the SG achieved a better, statistically significant result than the CG in the final test both in 2017–2018 and 2018–2019, even though the situation at the beginning of the physics course was reversed in the first year and equivalent in the second one.

Moving to RQ4, although the ITM appeared to be more successful than the traditional teaching strategy regardless of the level of difficulty of the topics tested, the intensity of the effect was considerably different between the two academic years, i.e., it was decidedly lower when the test was more difficult (lower mean DI). However, the physics knowledge of the CG students who took both the initial and final tests in 2018–2019, when the level of difficulty was higher, did not show statistically significant differences between initial and final test. On the one hand, the size of the improvement in physics conceptual understanding that characterized the SG freshmen who took both the tests was negligible in 2018–2019. On the other hand, the traditional methods were not able to definitely determine any improvements in the physics learning of the CG freshmen.

All in all, according to the scientific literature, PL might help to improve students' physics learning more than traditional lectures but the novelty in this study is that it may have appreciable, above average results in university LSC. In spite of the difficulties analyzing the role played by the instructors, the ITM was found to be consistently advantageous in several analyses carried out to test its effect. This is certainly an issue that requires further investigation when dealing with PL in LSC.

## 5. Conclusions

Physics educational research has widely investigated the effectiveness of employing many different active methods in academic introductory physics courses. Many studies generally indicate that these teaching strategies increase class attendance and learners' engagement and improve students' conceptual understanding of the physics topics, albeit to varying degrees depending on the specific methodology adopted. Conversely, there is not much agreement on such strategies' impact on the acquisition of other important skills by the learners, for instance problem solving, in comparison with a traditional approach. Moreover, many of the studies have been carried out in the context of small or medium size classes, whilst academic institutions face the increasing challenge of the massification of higher education resulting in LSC.

Addressing this limitation in the scientific literature, in this study we analyzed the problem of how to improve university students' physics learning from a novel perspective, investigating the effectiveness of an educational strategy based on the synergetic integration of PL activities, strengthened by the use of technology, into traditional physics lectures and drills as a physics teaching method in university LSC. As a consequence, the challenge of adopting this integrated strategy is to value the merits of both active methods and traditional lectures and drills, and limit their own flaws. All in all, this combined strategy has proved to be effective in fostering learners' conceptual understanding of physics in the context of LSC and more successful than traditional courses centered on classical lectures. Although the efficacy of the methodology was confirmed for physics topics characterized by dissimilar levels of difficulty, the size of the effect tended to be appreciably different: it was large when the topics tested were easier, but small if they were more complex.

However, if a correlation can be argued between the employment of this teaching strategy and the improvement of the students' level of knowledge in physics, the instructor's role and their contribution to the undergraduates' learning will be further investigated.

Furthermore, a threshold for the exposure to PL in LSC in order for it to be effective has been highlighted. This minimum level of exposure appears to depend on the complexity of the physics topics addressed, but at any rate it corresponds to a small percentage (2–2.5%) of the total time of the physics course attended by freshmen.

This is an important result considering the common reasons frequently indicated by instructors to explain their resistance to active learning strategies, such as lack of time. A possible transition from a traditional approach to this ITM would not require an instructor to revolutionize their physics course design and their approach to teaching. Moreover, implications for STEM teaching also emerge from the present study: this ITM may be applied to a wide range of scientific disciplines and courses with minor changes, i.e., adapting the tests and the PL activities to the specific subjects.

These promising results may thus foster ongoing changes in university policy towards the renewal of the teaching methodology, which is actually urged by policy makers in the European Union and the United States, among others.

Finally, further studies should explore possible improvements in students learning with reference to a higher level of exposure to PL activities and a more intensive use of technology.

**Author Contributions:** Conceptualization, M.B., J.E.R. and M.Z.; Data curation, M.B.; Formal analysis, M.B.; J.E.R. and M.Z.; Investigation, M.Z.; Methodology, J.E.R. and M.Z.; Project administration, M.Z.; Resources, M.Z.; Software, M.B., J.E.R.; Supervision, M.Z.; Visualization, M.Z.; Writing—original draft, M.B. 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 according to the guidelines of the Declaration of Helsinki, and approved by the Data Protection Controller.

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

**Data Availability Statement:** The data presented in this study are openly available in M. Bozzi, J. Raffaghelli and M. Zani, Datasets from the study Peer learning as a key component of an Integrated Teaching Method in large size classes [Data set]. Zenodo, (2020). <http://doi.org/10.5281/zenodo.3957433> (accessed on 27 December 2020).

**Acknowledgments:** The authors wish to thank Universitat Oberta de Catalunya for supporting the inferential analysis development and Claudia Andreani for editing this paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Angell, C.; Guttersrud, Ø.; Henriksen, E.K.; Isnes, A. Physics: Frightful, but fun. Pupils' and teachers' views of physics and physics teaching. *Sci. Educ.* **2004**, *88*, 683–706. [[CrossRef](#)]
2. Örnek, F.; Robinson, W.R.; Haugan, M.R. What Makes Physics Difficult? *Sci. Educ. Int.* **2007**, *18*, 165–172.
3. Redish, E.F.; Scherr, R.E.; Tuminaro, J. Reverse-Engineering the Solution of a "Simple" Physics Problem: Why Learning Physics Is Harder Than It Looks. *Phys. Teach.* **2006**, *44*, 293–300. [[CrossRef](#)]
4. Bozzi, M.; Balossi, B.; Di Sieno, L.; Ganzer, L. Securing freshmen's learning through a Physics refresher course: A breakthrough experience at Politecnico di Milano. In Proceedings of the XII International Conference of Education, Research and Innovation (iCERi2019), Seville, Spain, 11–13 November 2019; pp. 2237–2243.
5. Stains, M.; Harshman, J.; Barker, M.K.; Chasteen, S.V.; Cole, R.; DeChenne-Peters, S.E.; Eagan, M.K., Jr.; Esson, J.M.; Knight, J.K.; Laski, F.A. Anatomy of STEM teaching in North American universities. *Science* **2018**, *359*, 1468–1470. [[CrossRef](#)] [[PubMed](#)]
6. King, A. From Sage on the Stage to Guide on the Side. *Coll. Teach.* **1993**, *41*, 30–35. [[CrossRef](#)]
7. Bransford, J.D.; Johnson, M.K. Contextual Prerequisites for Understanding: Some Investigations of Comprehension and Recall. *J. Verb. Lear. Verb. Behav.* **1972**, *11*, 717–726. [[CrossRef](#)]
8. Hammer, D.; Elby, A. Tapping Epistemological Resources for Learning Physics. *J. Learn. Sci.* **2003**, *12*, 53–90. [[CrossRef](#)]
9. Redish, E.F. The implications of cognitive studies for teaching physics. *Am. J. Phys.* **1994**, *62*, 796–803. [[CrossRef](#)]

10. European Commission. *High Level Group on the Modernisation of Higher Education*; Report to the European Commission on Improving the Quality of Teaching and Learning in Europe's Higher Education Institutions, Publications Office of the European Union: Luxembourg, 2013.
11. Li, T.; Yang, N. Comparing MOOCs with Traditional Courses for Quality Teaching in Higher Education. In Proceedings of the 2nd International Conference on Modern Education and Information Technology (MEIT 2018), Xiamen, China, 23–24 June 2018; DEStech Publications, Inc.: Lancaster, PA, USA, 2018; pp. 122–127.
12. Duch, B.J. Problem-based learning in physics: The power of students teaching students. *JCST* **1996**, *15*, 326–329.
13. Hake, R.R. Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *Am. J. Phys.* **1998**, *66*, 64–74. [[CrossRef](#)]
14. Vygotsky, L.S. *Thought and Language*; Hanfmann, E., Vakar, G., Eds.; MIT Press: Cambridge, MA, USA, 1962.
15. Crouch, C.; Mazur, E. Peer Instruction: Ten years of experience and results. *Am. J. Phys.* **2001**, *69*, 970–977. [[CrossRef](#)]
16. Fraser, J.M.; Timan, A.L.; Miller, K.; Dowd, J.E.; Tucker, L.; Mazur, E. Teaching and physics education research: Bridging the gap. *Rep. Prog. Phys.* **2014**, *77*, 1–17. [[CrossRef](#)] [[PubMed](#)]
17. Freeman, S.; Eddy, S.L.; McDonough, M.; Smith, M.K.; Okoroafor, N.; Jordt, H.; Wenderoth, M.P. Active learning increases student performance in science, engineering, and mathematics. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 8410–8415. [[CrossRef](#)]
18. Adams, W.K.; Perkins, K.K.; Podolefsky, N.S.; Dubson, M.; Finkelstein, N.D.; Wieman, C.E. New instrument for measuring student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey. *Phys. Rev. ST Phys. Educ. Res.* **2006**, *2*, 1–14. [[CrossRef](#)]
19. Brewster, E.; Kramer, L.H.; O'Brien, G. Modeling instruction: Positive attitudinal shifts in introductory physics measured with CLASS. *Phys. Rev. ST Phys. Educ. Res.* **2009**, *5*, 1–5. [[CrossRef](#)]
20. Watkins, J.; Mazur, E. Retaining students in science, technology, engineering, and mathematics (STEM) majors. *JCST* **2013**, *42*, 36–41.
21. Hardy, J.; Hancock, S.; Johnson, O.; Galloway, R.; Wallace, M.; Draper, S.; Bates, S. *PER-Based Reform*; Redish, E.F., Cooney, P.J., Eds.; American Institute of Physics: College Park, MD, USA, 2007; pp. 1–42. Available online: <http://www.per-central.org/document/ServeFile.cfm?ID=4517> (accessed on 20 August 2020).
22. European Commission/EACEA/Eurydice. *The European Higher Education Area in 2015: Bologna Process Implementation Report*; Publications Office of the European Union: Luxembourg, 2015.
23. Ghislandi, P.; Raffaghelli, J. Quality teaching matters: Perspectives on quality teaching for the modernization of higher education. A position paper. *Form. Insegn.* **2014**, *XII*, 57–81.
24. Ranieri, M.; Raffaghelli, J.E.; Bruni, I. Supporting Learning Design as a driver for pedagogical innovation within an integrated model of faculty development. In *Faculty Development for Digital Teaching and Learning*; Elçi, A., Beith, L.L., Elçi, A., Eds.; IGI Global: Hershey, PA, USA, 2019; pp. 77–98.
25. Deslauriers, L.; McCarty, L.S.; Miller, K.; Callaghan, K.; Kestin, G. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 19251–19257. [[CrossRef](#)] [[PubMed](#)]
26. Silverthorn, D.U.; Thorn, P.M.; Svinicki, M.D. It's difficult to change the way we teach: Lessons from the Integrative Themes in Physiology curriculum module project. *Adv. Physiol. Educ.* **2006**, *30*, 204–214. [[CrossRef](#)] [[PubMed](#)]
27. McDaniel, M.A.; Stoen, S.; Frey, R.F.; Markow, Z.E.; Hynes, K.M.; Zhao, J.; Cahill, M.J. Dissociative conceptual and quantitative problem solving outcomes across interactive engagement and traditional format introductory physics. *Phys. Rev. Phys. Educ. Res.* **2016**, *12*, 1–16. [[CrossRef](#)]
28. Schiltz, G.; Feldman, G.; Vaterlaus, A. Active-learning settings and physics lectures: A performance analysis. *J. Phys. Conf. Ser.* **2019**, *1286*, 1–8. [[CrossRef](#)]
29. Hornsby, D.J.; Osman, R. Massification in higher education: Large classes and student learning. *High. Educ.* **2014**, *67*, 711–719. [[CrossRef](#)]
30. Palfreyman, D.; Tapper, T.; Thomas, S. Series editor's introduction: International studies in higher education. In *Accountability in Higher Education. Global Perspectives on Trust and Power*; Stensaker, B., Harvey, L., Eds.; Routledge: London, UK, 2011; pp. 13–15.
31. Scott, P. *The Meanings of Mass Higher Education*; Open University Press: Buckingham, UK, 1995.
32. Trees, A.R.; Jackson, M.H. The learning environment in clicker classrooms: Student processes of learning and involvement in large university-level courses using student response systems. *Learn. Media Tech.* **2007**, *32*, 21–40. [[CrossRef](#)]
33. Deslauriers, L.; Schelew, E.; Wieman, C. Improved Learning in a Large-Enrollment Physics Class. *Science* **2011**, *332*, 862–864. [[CrossRef](#)] [[PubMed](#)]
34. Henderson, C.; Dancy, M.H. Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics. *Phys. Rev. ST Phys. Educ. Res.* **2007**, *3*, 1–14. [[CrossRef](#)]
35. Leinonen, R.; Asikainen, M.A.; Hirvonen, P.E. Peer discussions in lecture-based tutorials in introductory physics. *Phys. Rev. Phys. Educ. Res.* **2017**, *13*, 1–13. [[CrossRef](#)]
36. Wood, A.K.; Galloway, R.K.; Hardy, J.; Sinclair, C.M. Analysing learning during Peer Instruction dialogues: A resource activation framework. *Phys. Rev. Phys. Educ. Res.* **2014**, *10*, 1–15.
37. Pellegrini, M. The effectiveness of peer assessment on students' performance in higher education. Evidence of an overview of meta-analyses. *Form@re* **2020**, *20*, 128–142.

38. Raffaghelli, J.; Ghislandi, P.; Sancassani, S.; Canal, L.; Micciolo, R.; Balossi, B.; Bozzi, M.; Di Sieno, L.; Genco, I.; Gondoni, P.; et al. Integrating MOOCs in physics preliminary undergraduate education: Beyond large size lectures. *Educ. Media Int.* **2018**, *55*, 301–316. [CrossRef]
39. Gupta, A.; Hammer, D.; Redish, E.F. The Case for Dynamic Models of Learners' Ontologies in Physics. *J. Learn. Sci.* **2010**, *19*, 285–321. [CrossRef]
40. Ausubel, D.P. *The Psychology of Meaningful Verbal Learning*; Grune and Stratton: New York, NY, USA, 1963.
41. Ausubel, D.P. *Educational Psychology: A Cognitive View*; Rinehart, & Winston: Holt, MO, USA, 1968.
42. Ausubel, D.P. *The Acquisition and Retention of Knowledge*; Kluwer: Dordrecht, The Netherlands, 2000.
43. Novak, J.D. Meaningful Learning: The Essential Factor for Conceptual Change in Limited or Inappropriate Propositional Hierarchies Leading to Empowerment of Learners. *Sci. Educ.* **2002**, *86*, 548–571. [CrossRef]
44. Vygotsky, L.S. Interaction between learning and development. In *Mind and Society—The Development of Higher Psychological Processes*; Vygotsky, L.S., Cole, M., Eds.; Harvard University Press: Cambridge, MA, USA, 1978; pp. 79–91.
45. Vygotsky, L.S.; Wertsch, J.F. *The Concept of Activity in Soviet Psychology*; Sharpe: Armonk, NY, USA, 1981; pp. 134–240.
46. Marton, F.; Säljö, R. On qualitative differences in learning: I—Outcome and process. *Br. J. Educ. Psychol.* **1976**, *46*, 4–11. [CrossRef]
47. Biggs, J. *Teaching for Quality Learning at University*; SHRE and Open University Press: Maidenhead, Berkshire, UK, 1999.
48. Ramsden, P. *Learning to Teach in Higher Education*; Routledge: London, UK; New York, NY, USA, 1992.
49. American Association for the Advancement of Science. *Vision and Change: A Call to Action. Final Report*; American Association for the Advancement of Science: Washington, DC, USA, 2011.
50. President's Council of Advisors on Science and Technology. *Engage to Excel: Producing one Million Additional College Graduates with Degrees in Science*; Government Office of Science and Technology: Washington, DC, USA, 2012.
51. European Commission. *Does the EU need more STEM Graduates? Final Report*; Publications Office of the European Union: Luxembourg, 2016.
52. DM (Ministerial Decree) 29 December 2014 n. 976. Available online: <http://attinisteriali.miur.it/anno-2014/dicembre/dm-29122014.aspx> (accessed on 27 December 2020).
53. DM (Ministerial Decree) 29 December 2017 n. 1047. Available online: <https://www.miur.gov.it/documents/20182/226551/DM+1047+del+29+dic+2017.pdf/6d585918-e820-4966-9c13-d6790f30fb3a?version=1.1> (accessed on 27 December 2020).
54. Zani, M.; Bozzi, M. La fisica tra la scuola secondaria e l'università. Riflessioni e orientamenti [Physics between secondary school and university. Reflections and orientations]. *Nuova Second.* **2018**, *XXXVI*, 84–88.
55. Prince, M. Does Active Learning Work? A Review of the Research. *J. Eng. Educ.* **2004**, *93*, 223–231. [CrossRef]
56. Ballen, C.J.; Wieman, C.; Salehi, S.; Searle, J.B.; Zamudio, R.K. Enhancing Diversity in Undergraduate Science: Self-Efficacy Drives Performance Gains with Active Learning, CBE. *Life Sci. Educ.* **2017**, *16*, 56. [CrossRef]
57. Bradforth, S.E.; Miller, E.R.; Dichtel, W.R.; Leibovich, A.K.; Feig, A.L.; Martin, J.D.; Bjorkman, K.S.; Schultz, Z.D.; Smith, T.L. University learning: Improve undergraduate science education. *Nat. Cell Biol.* **2015**, *523*, 282–284. [CrossRef] [PubMed]
58. Hattie, J. *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*; Routledge: London, UK, 2008.
59. Hattie, J. The applicability of Visible Learning to higher education. *Scholarsh. Teach. Learn. Psychol.* **2015**, *1*, 79–91. [CrossRef]
60. Yang, N.; Ghislandi, P.; Dellantonio, S. Online collaboration in a large university class supports quality teaching. *Educ. Tech. Res. Dev.* **2018**, *66*, 671–691. [CrossRef]
61. Bowman-Perrott, L.; Davis, H.; Vannest, K.; Williams, L.; Greenwood, C.; Parker, R. Academic Benefits of Peer Tutoring: A Meta-Analytic Review of Single-Case Research. *Sch. Psychol. Rev.* **2013**, *42*, 39–55. [CrossRef]
62. Leung, K.C. A preliminary empirical model of crucial determinants of best practice for peer tutoring on academic achievement. *J. Educ. Psychol.* **2015**, *107*, 558–579. [CrossRef]
63. Griffiths, S.; Houston, K.; Lazenbatt, A. *Enhancing Student Learning Through Peer Tutoring in Higher Education*; Educational Development Unit, University of Ulster: Coleraine, UK, 1995.
64. Boud, D. Moving towards autonomy. In *Developing Student Autonomy in Learning*, 2nd ed.; Boud, D., Ed.; Kogan Page: London, UK, 1988.
65. Boud, D. Making the Move to Peer Learning. In *Peer Learning in Higher Education: Learning From and With Each Other*; Boud, D., Cohen, D.R., Sampson, J., Eds.; Kogan Page: London, UK, 2001; pp. 1–20.
66. Damon, W.; Phelps, E. Critical distinctions among three approaches to peer education. *Int. J. Educ. Res.* **1989**, *13*, 9–19. [CrossRef]
67. Brooks, D.C. Space matters: The impact of formal learning environments on student learning. *Br. J. Educ. Technol.* **2011**, *42*, 719–726. [CrossRef]
68. Cotner, S.; Loper, J.; Walker, J.D.; Brooks, D.C. "It's Not You, It's the Room"—Are the High-Tech, Active Learning Classrooms Worth It? *JCST* **2013**, *42*, 82–88.
69. Jensen, J.L.; Kummer, T.A.; Godoy, P.D.D.M. Improvements from a Flipped Classroom May Simply Be the Fruits of Active Learning. *CBE Life Sci. Educ.* **2015**, *14*, 5. [CrossRef] [PubMed]
70. Stoltzfus, J.R.; Libarkin, J. Does the Room Matter? Active Learning in Traditional and Enhanced Lecture Spaces. *CBE Life Sci. Educ.* **2016**, *15*, 68. [CrossRef] [PubMed]
71. Krueger, A.B. Economic Considerations and Class Size. *Econ. J.* **2003**, *113*, 34–63. [CrossRef]
72. Mitchell, M.; Palacios, V.; Leachman, M. States Are Still Funding Higher Education Below Pre-Recession Levels. *JCBA* **2015**, *10*, 71.

73. Estermann, T.; Pruvot, E.B.; Claeys-Kulik, A.L. *Designing Strategies for Efficient Funding of Higher Education in Europe. DEFINE Report*; European University Association: Brussels, Belgium, 2013.
74. Kokkelenberg, E.; Dillon, M.; Christy, S. The effects of class size on student grades at a public university. *Econ. Educ. Rev.* **2008**, *27*, 221–233. [[CrossRef](#)]
75. Geske, J. Overcoming the Drawbacks of the Large Lecture Class. *Coll. Teach.* **1992**, *40*, 151–154. [[CrossRef](#)]
76. Hamilton, W.R.; Padron, V.A.; Henriksen, J.A. Engaging Students in a Large Classroom and Distance Environment. In *Handbook of Research on Teaching and Learning in K-20 Education*; Wang, V.X., Ed.; IGI Global: Hershey, PA, USA, 2013; pp. 759–777.
77. Ranieri, M.; Raffaghelli, J.E.; Bruni, I. Game-based student response system: Revisiting its potentials and criticalities in large-size classes. *Act. Learn. High. Educ.* **2018**. [[CrossRef](#)]
78. Maddison, T. A Matter of Size: Flipping Library Instruction in Various Engineering Classrooms. *Issues Sci. Technol. Librariansh.* **2015**, *82*, 1–9.
79. Martín-SanJosé, J.-F.; Juan, M.-C.; Seguí, I.; García-García, I. The effects of computer-based games and collaboration in large groups vs. collaboration in pairs or traditional methods. *Comput. Educ.* **2015**, *87*, 42–54. [[CrossRef](#)]
80. Nicol, D.J.; Boyle, J.T. Peer Instruction versus Class-wide Discussion in Large Classes: A comparison of two interaction methods in the wired classroom. *Stud. High. Educ.* **2003**, *28*, 457–473. [[CrossRef](#)]
81. Dufresne, R.J.; Gerace, W.J.; Leonard, W.J.; Mestre, J.P.; Wenk, L. Classtalk: A classroom communication system for active learning. *J. Comput. High. Educ.* **1996**, *7*, 3–47. [[CrossRef](#)]
82. McDermott, L.C.; Shaffer, P.S. *Tutorials in Introductory Physics*; Pearson: Boston, MA, USA, 2010.
83. Sands, D.; Parker, M.; Hedgeland, H.; Jordan, S.; Galloway, R. Using concept inventories to measure understanding. *High. Educ. Pedagog.* **2018**, *3*, 173–182. [[CrossRef](#)]
84. Hestenes, D.; Wells, M.; Swackhamer, G. Force Concept Inventory. *Phys. Teach.* **1992**, *30*, 141–158. [[CrossRef](#)]
85. Cook, T.D.; Campbell, D.T. The causal assumptions of quasi-experimental practice—The origins of quasi-experimental practice. *Synthese* **1986**, *68*, 141–180.
86. Montero, I.; León, O.G. A guide for naming research studies in Psychology. *Int. J. Clin. Health Psychol.* **2007**, *7*, 847–862.
87. Cohen, L.; Manion, L.; Morrison, K. *Research Methods in Education*, 7th ed.; Routledge: London, UK; New York, NY, USA, 2011.
88. Hammersley, M. Methodological Paradigms in Educational Research. 2012. Available online: <https://martynhammersley.files.wordpress.com/2013/11/methodological-paradigms-4.pdf> (accessed on 27 December 2020).
89. McDermott, L.C.; Shaffer, P.S. Research as a guide for curriculum development: An example from introductory electricity. Part I: Investigation of student understanding. *Am. J. Phys.* **1992**, *60*, 994–1003. [[CrossRef](#)]
90. Hammer, D. Student resources for learning introductory physics. *Am. J. Phys.* **2000**, *68*, 52–59. [[CrossRef](#)]
91. Ebel, R.L.; Frisbie, D.A. *Essentials of Educational Measurement*; Prentice-Hall: Englewood Cliffs, NJ, USA, 1986.
92. Crocker, L.; Algina, J. *Introduction to Classical and Modern Test Theory*; Holt Rinehart and Winston: New York, NY, USA, 1986.
93. Guarascio, A.J.; Nemecek, B.D.; Zimmerman, D.E. Evaluation of students' perceptions of the Socratic application versus a traditional student response system and its impact on classroom engagement. *Curr. Pharm. Teach. Learn.* **2017**, *9*, 808–812. [[CrossRef](#)] [[PubMed](#)]
94. Lim, W.N. Improving Student Engagement in Higher Education through Mobile-Based Interactive Teaching Model Using Socratic. In Proceedings of the 8th IEEE Global Engineering Education Conference (EDUCON 2017), Athens, Greece, 25–28 April 2017; pp. 404–412.
95. Afreen, R. Bring Your Own Device (BYOD) in Higher Education: Opportunities and Challenges. *IJETTCS* **2014**, *3*, 233–236.
96. Blasco-Arcas, L.; Buil, I.; Hernández-Ortega, B.; Sese, F.J. Using clickers in class. The role of interactivity, active collaborative learning and engagement in learning performance. *Comput. Educ.* **2013**, *62*, 102–110. [[CrossRef](#)]
97. Caldwell, J.E. Clickers in the Large Classroom: Current Research and Best-Practice Tips. *CBE Life Sci. Educ.* **2007**, *6*, 9–20. [[CrossRef](#)] [[PubMed](#)]
98. Mayer, R.E.; Stull, A.; DeLeeuw, K.; Almeroth, K.; Bimber, B.; Chun, D.; Bulger, M.; Campbell, J.; Knight, A.; Zhang, H. Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes. *Contemp. Educ. Psychol.* **2009**, *34*, 51–57. [[CrossRef](#)]
99. Bozzi, M.; Raffaghelli, J.; Zani, M. Datasets from the study Peer learning as a key component of an Integrated Teaching Method in large size classes [Data set]. *Zenodo* **2020**. [[CrossRef](#)]
100. Sawilowsky, S.S. New Effect Size Rules of Thumb. *J. Mod. Appl. Stat. Methods* **2009**, *8*, 597–599. [[CrossRef](#)]