Learning Newtonian Physics through Programming Robot Experiments

Paola Ferrarelli¹ · Luca Iocchi¹

Accepted: 26 February 2021 / Published online: 18 March 2021
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract

Novel technology has been applied to improve students' learning abilities in different disciplines. The research in this field is still finding suitable methodologies, tools, and evaluation mechanisms to devise learning frameworks with high impact on students' performance. This article describes an instructional method to perform Newtonian physics experiments by programming a mobile robot. Such experiments allow the learners to design, implement and visualize physics concepts, thus using the robot as a *cognitive tool* or *mindtool*. An accurate assessment of the students learning gain, involving 29 high-school students, shows that the proposed method provided significant improvements in the students understanding of the first Newton's law, the second Newton law and the superposition principle. The learning gain has been measured through the Force Concept Inventory questionnaire. From this study, we can state that programming a mobile robot to perform physics experiments can improve knowledge about Newtonian physics, even without giving specific lectures in the subject and with a much shorter lecture plan with respect to traditional lectures.

1 Introduction

Technology is an ever-evolving process affecting and accompanying contemporary industrial, social and educational progress. Consequently, the demand for an information society makes the use of advanced technological tools increasingly widespread in education, but, at the same time, usually not present in the school curriculum. Indeed, in recent years, we have witnessed an impressive increase in applications using robotics and artificial intelligence. These technologies are now mature enough to be effectively combined with educational activities, engaging with students in new ways and helping teachers to teach more effectively at school Timms (2016).

We believe that robotics and artificial intelligence can be applied to many subjects, given its general applicability. When applied to school curricular subjects, these

✉ Paola Ferrarelli
ferrarelli@diag.uniroma1.it

Luca Iocchi
iocchi@diag.uniroma1.it

¹ Department of Computer, Control, and Management Engineering, Sapienza University of Rome, Rome, Italy

disciplines have the advantage of being easier to be accepted by teachers who can even exploit these technologies to refine and reshape their learning goals and methods.

On the other hand, the introduction of new technologies in education should not only expose students to such technologies for providing better opportunities of choosing future studies and jobs, but also improve the learning process in curricular subjects. The idea of using digital technologies to support the development of mental functioning has its roots in Vygotsky social constructivist psychology, according to which educational technologies are *cognitive tools* that work as partners in learning Vygotsky (2012).

From a research perspective, the creation and use of new cognitive tools at school enable the study of changes in mental processes and outcomes of human learning. In this respect, by the mid-1980s, Pea (1985) highlighted the difference between technological tools used for increasing efficiency in achieving a task, and technological tools used to reorganize the mental schemata of learners. The latter were defined as *educational* or *cognitive* tools. For Salomon et al. (1991) cognitive tools can help learners to perform tasks at a higher level of cognition. They argued that the desired educative effects of cognitive tools cannot be expected automatically but must be designed into the tools and the contexts in which they are used.

By the mid-1990s, Jonassen and colleagues became the principal proponents of the concept of educational technology as cognitive tools or *mindtools* that help learners to effectively model, organize, visualize, and interpret their knowledge. There are several kinds of mindtools, including semantic organization tools, dynamic modelling tools, information interpretation tools, knowledge construction tools, and conversation and collaboration tools. In these studies, cognitive tools showed the potential to open a path to constructivist learning and transferable knowledge in educational systems dominated by routinized practice and reproduction of given information Jonassen (1996). Jonassen also described how students' intentional engagement with cognitive tools promotes meaningful learning and reflective thinking Jonassen (2000).

Many other studies demonstrated that students could learn more effectively when they actively participate in discussion and experimentation in the classroom.

For example, Hake (1998) investigated the understanding of the Newtonian concepts of the force of 6000 students, across a wide variety of institutions, applying different teaching methodologies: traditional lecture-based lessons versus interactive instructional lessons. Using the Force Concept Inventory test (FCI), developed by Hestenes et al. (1992), he found significant improvements in the students learning gains after interactive engagement lessons.

Von Korff et al. (2016) confirmed Hake results with a much larger sample representing about 450 classes and about 31,000 students. They analyzed the FCI data published between 1995 and 2014: the interactive teaching techniques are significantly more likely to produce high student learning gains than traditional lecture-based instruction. They also established that interactive instruction works in many settings, including students with a high and low level of prior knowledge, at liberal arts and research universities, and enrolled in small and large classes.

Following all these research streams, this article presents an instructional method that empowers and assesses learning of Newtonian physics curricular concepts through programming a mobile robot to perform relevant experiments in the subject. In our work, the robot acts as a cognitive tool accompanying the students to accomplish their tasks (i.e., experiments in Newtonian physics). In particular, the robot acts as a dynamic modelling tool or a knowledge construction tool, as defined by Jonassen et al. (1998).

More specifically, we designed and developed an educational project (i.e., a set of educational activities), called *Lab2GO-Robotica*, targeting high school students (age 15–19) from different schools who worked in small teams, under the guidance of a teacher, to build and program a mobile robot. During the project, 29 students performed experiments about Newtonian physics with the robots in a highly interactive environment, with a continuous exchange of ideas between students and teachers and among students themselves.

Building on Pea's and Jonassen's ideas, our conjecture is that programming a mobile robot to perform specific experiments in a subject has the potential to change how cognition is organized. The research question we want to address in our work is

To what extent programming a mobile robot to perform physics experiments can increase high school student's knowledge of Newtonian physics concepts?

To detect and measure cognitive changes due to the educational activity, we defined an experimental protocol to assess the performance of an experimental group in terms of learning gain in physics concepts. As for performance measure, we used the FCI test, developed by Hestenes et al. (1992), that has already been validated and successfully used by the physics education research community to assess the performance of innovative educational tools and courses in physics. However, to the best of our knowledge, no reports evaluating the performance of using robots to learn physics, based on commonly accepted measures, are available.

The contributions of this article are thus: (1) the development of an educational project for teaching students how to program a mobile robot to perform experiments relevant to Newtonian physics; (2) an experimental protocol to evaluate the effectiveness of such a project; (3) the analysis of the results highlighting the impact of the educational robotics activities to Newtonian physics concepts.

The article is organized as follows: Sect. 2 presents related work and highlights the novelty of our project; Sect. 3 summarizes the main features of the robotic platform used in the project; Sect. 4 illustrates the design of the instructional method; while Sect. 5 describes its implementation. The results are discussed in Sects. 6, 7, and 8. Finally, Sect. 9 introduces some important lessons learned during the development of our project and Sect. 10 draws some conclusions.

2 Related Work

2.1 Instructional Methods Using Robots

A survey on the use of robots in the educational field was made by Benitti (2012) and Mubin et al. (2013).

Benitti reviewed the international literature on educational robotics (ER) published over 10 years, listing used robots, student's age, obtained results, and research issues; Mubin et al. (2013) instead gave an overview of robot kits, robot roles and robot usage domains.

From Benitti literature analysis, two relevant issues emerged: (1) most of the educational activities with robots targeted subjects that were closely related to robotics (such as robot building and programming) and were performed during post-school programs, workshops or summer camps, rather than at school; (2) only a few studies contained a quantitative evaluation of their results, through statistical methods. Indeed, the evaluation methodology should clearly distinguish the experimental group (i.e., students

involved in the program under evaluation) and the control group (i.e., students participating to other activities) to compare the results obtained by the two groups. An essential step of this methodology is the selection of the participants in each group. According to Benitti (2012), only 44% of studies had a quantitative assessment of learning. Within these works, only 60% used a control group in the experimental protocol and only 20% adopted a random selection to recruit students in the two groups.

As for the first point, we believe that a better alignment of the ER activities with the students' school curriculum might improve the acceptability of the teachers' educational variations and the expenses (in terms of time and money) by the school managers. For this reason, we decided to use robotics to explore Newtonian physics concepts, that are part of the physics high school curriculum. We expected to obtain learning improvements, as the research conducted in physics education shows that traditional physics courses, consisting of lectures, generate a low level of learning even if taught by very good and expert professors Hestenes et al. (1992).

As for the second point, we believe that a suitable assessment of the students learning after the ER activities is fundamental to progress the research in the field and to justify the investment made by the school. We decided to assess our methodology using quantitative data, treated statistically. We also compared the performance of the experimental group of students with the performance of a comparison group formed by their classmates, who did not use robots to review Newtonian physics concepts. The results obtained so far and discussed in this article provide a clear picture of the students' learning gain involved in this experimentation.

Regarding the study of robots for learning physics concepts, a literature review of instructional strategies to promote the conceptual change in students' thinking about force and motion in physics was done by Tomara et al. (2017). Still, none of the described approaches included the use of a mobile robot. On the other hand, when robots were used for educational activities related to physics concepts, no comprehensive, systematic and quantitative results, based on consolidated and validated performance measures, were reported. For example, Church et al. (2010) described the successful use of Lego® Mindstorms® robot kit in designing robotics-based activities for teaching high school physics classes: they observed great student engagement. Alimisis (2012) highlighted the role of constructivist pedagogy while using robotics to teach kinematics concepts in physics. Mitnik et al. (2008) presented the application of robotic technologies to teach distances, angles, kinematics, graph construction, and interpretation. They used ad-hoc questionnaires to assess the physics content knowledge. Finally, a robotics summer camp on physics content knowledge of middle school students' was evaluated by Williams et al. (2007) with an ad-hoc questionnaire.

In contrast with the above-cited work using ad-hoc evaluation protocols, the project Lab2GO-Robotica described in this article has been evaluated with the FCI test, developed by Hestenes et al. (1992), that has already been validated and successfully used by the physics education research community.

In summary, the ER activity proposed in this article fits the school curriculum (Newtonian physics), is performed within the school year, and is evaluated with the FCI test, a performance measure of Newtonian physics concepts accepted by the physics education research community.

2.2 Educational Robots

Regarding the choice of the robotic platform, we believe that it must have features well-suited for the learning objectives and the high school students' abilities.

Several educational robots are available with different features and capabilities. For example, ROBIN is an agent learning companion that provides cognitive and social feedback on the students' programming task to move a Lego®Mindstorms® robot Ahmed et al. (2018). Programming a small mobile robot with a sequence of commands has been used to illustrate concepts of Mathematics and Geometry to very young children (pre-school and primary school) Ferrarelli et al. (2017).

Anthropomorphic or zoomorphic robots, such as SoftBank Robotics Nao or Innvo Labs Pleo the dinosaur can provide social interactions, talk, show facial expressions, etc. and are mostly used to teach non-technical subjects, such as English verbs Tanaka and Matsuzoe (2012). The development of a motivating, interactive learning environment, featuring a social robot in computer science, was explored by Pfeifer and Lugin (2018). A study to investigate the increase of the student's curiosity on the science used the Robovie social robot, that has behaviours designed to initiate a conversation about science topics and answer any questions related to it, in a semi-autonomous mode, acting as a peer with 4th-5th grade students Shiomi et al. (2015).

Finally, Hussain et al. (2006) demonstrated that there is not a generally positive attitude towards Lego®Mindstorms® material among the pupils. He assessed that only grade 5 and grade 9 students with higher ability in mathematics tend to be more engaged by Lego®Mindstorms® material than others. Robots that need to be assembled before being used or that are complex to use could not fit well for young students.

In this article, we describe the robotic platform MARRtino developed in our University to integrate robotics and artificial intelligence techniques with teaching materials and a comfortable software layer to use by high school students. MARRtino robot uses state-of-the-art robotics components and is based on a standard development environment and can be easily extended to integrate new functionalities. Nonetheless, as described in Sect. 3, we implemented a software layer that makes the robot easy-to-use as other educational platforms.

The main advantages of using MARRtino are its full hardware and software open-source design and implementation and the use of standard development tools that are well-known and largely used by the robotics research and industrial communities.

3 MARRtino Educational Robot

The Lab2GO-Robotica project has been developed using the MARRtino mobile robot, a hardware and software open-source platform integrating artificial intelligence and robotics technologies and educational material that explains how to build and program the robot, using simple programming interfaces.

MARRtino¹ is a simple differential drive robot composed by the following main components: a box chassis made of Plexiglas, two motors with encoders, an Arduino Mega 2560 running the firmware to control the motors, a Motor Shield to drive the two motors,

¹ <https://www.marrtino.org/>.

a 12V battery, a LED display to show the battery voltage, a switch to turn on/off the robot, a charger, and a Raspberry Pi to run high-level programs. Additional equipment, such as speakers, microphone, cameras, lasers and possibly other sensors can be attached to the USB ports of the Raspberry Pi to increase its functionalities.

MARRtino is a robot based on the Robot Operating System (ROS)², thus exploiting all the available open-source components developed by the ROS community. In this framework, we have developed a ROS node running on the Raspberry Pi that connects through a serial protocol with Arduino Mega and allows control of robot wheels and reading of encoders through standard ROS topics (i.e., `cmd_vel`, `odom`). In this way, MARRtino is compatible with other similar ROS-based platforms (such as TurtleBot2,³ for example), but it is available at a much lower price and comes in a construction kit so that the students can build it.

Although ROS is essential in MARRtino project, as it guarantees the integration of state-of-the-art components with minimal effort, it is in general too challenging to learn for high-school students in a short time. Indeed, most of the students involved in the experiment described in this article were unfamiliar with any programming language before starting the project. Therefore, we developed a software layer to enable students to use ROS features without requiring them to learn its details. This layer has a back-end that allows the management of a ROS-based robot and a web-based front-end that enables users to manage ROS nodes running on the robot with a high-level user-friendly graphical user interface (MARRtino bring up web interface).⁴ With this web interface, it is possible to start the robot ROS nodes (either the real robot or the simulator in a Virtual Machine), sensor nodes (for cameras, lasers, and sonars), the audio server (for speech synthesis and speech understanding), navigation nodes (for mapping, localization, path planning, and obstacle avoidance), object recognition functionalities (based on pre-trained Convolutional Neural Networks (CNN) models), face detection, person detection, person tracking and person following, multi-modal interaction manager (MODIM) to interact with a touch screen (e.g., a tablet mounted on the robot). The web graphical user interface can be accessed from any device equipped with WiFi technology (like a tablet, smartphone, PC). MARRtino software running in the Raspberry Pi board mounted on the robot is configured to start a wireless access point and a web server at boot to allow users to directly connect to the robot without the need of any other network infrastructure or device.

Once ROS nodes and services are started, the robot can be programmed in different ways through high-level commands that activate specific behaviors of the robot. Table 1 summarizes such high-level commands in Python style. The students indeed learned to write Python programs with such simple high-level commands (e.g., `forward()`, `left()`, ...) and they could exercise most of them with a simulated robot based on Stage Vaughan (2008). The programming experience was made more user-friendly using Blockly, an open-source library that allows adding block-based visual programming to a web application Pasternak et al. (2017), Fraser (2015) (an example is shown in Fig. 1).

The possibility of using all these features makes MARRtino an exceptionally suited educational robot, where students can properly combine advanced artificial intelligence and robotics techniques (such as navigation in dynamic environments, object recognition through deep learning, speech recognition using cloud services, etc.) with a

² www.ros.org.

³ <https://www.turtlebot.com/turtlebot2/>.

⁴ More details in <https://www.marrtino.org/programming>.

Table 1 MARRtino high-level programming commands

# Motion	# Range distance
Forward(m)	d = obstacleDistance(deg=0)
Backward(m)	
Left(deg)	# Audio
Right(deg)	Sound(name)
Turn(deg)	Say(text, language='en')
SetSpeed(lx,az,time,stopend=False)	s = asr()
Stop()	
Goto(gx,gy,gth)	# Vision
GotoTarget(gx,gy)	Img = getImage()
p = getRobotPose()	Img = getWebImage(objcat=None)
	n = faceDetection(img)
# Utils	Label,conf = mobilenetObjrec(img)
wait(sec)	
b = marrtinoOK()	# modim
d = distance(p1,p2)	ShowImage(value, which='default')
	ShowText(value, which='default')

Fig. 1 Programming experience using Blockly

suitable setting considering their background. For example, students can quickly implement voice-controlled navigation, visual recognition, and interactive tasks. Additional examples of the use of MARRtino in educational competitions aiming at integrating

artificial intelligence and robotics techniques can be found in the material used to run the European RoboCup@Home Education Challenge 2019⁵ Tan et al. (2019).

4 Instructional Method - Design

As already mentioned, our research goal is to assess how programming a mobile robot to perform physics experiments improves students knowledge on this subject. Building on Salomon et al. (1991) idea that the desired educative effects of cognitive tools cannot be expected automatically but must be designed into the tools and the contexts in which they are used, we devised an instructional method for high school students, following the five first principles of instruction and their corollaries enunciated by Merrill (2002), which relate to creating learning environments.

In fact, Merrill reviewed instructional design theories to identify prescriptive principles that are common to various theories, such as “multiple approaches to understanding” by Gardner (2018), “collaborative problem solving” by Nelson (1999), “constructivist learning environments” by Jonassen and Rohrer-Murphy (1999), and “learning by doing” by Schank et al. (1999), among others.

The five first principles of instruction and their corollaries are:

1. *Problem-centered* learning is promoted when learners are engaged in solving real-world problems. Engagement is activated by showing the learner what he or she will be able to do at the end of the course (*showcase*), by involving them at operation and action level rather than at problem level only (*task level*), and finally through a progression of problem complexity (*problem progressions*).
2. *Activation Phase* learning is promoted when a relevant previous experience is activated. Learners are directed to recall and apply the previous experience (*previous experience*) to solve a new provided relevant experience (*new experience*) and are encouraged to organize the new knowledge using previous mental models, based on experience (*structure*). For Andre (1997) motivational themes into instruction can serve as an organizing structure, for example flying a space ship or playing golf;
3. *Demonstration Phase (show me)*: learning is promoted when the instruction demonstrates what is to be learned rather than merely telling. The demonstration must be consistent with the learning goal (*demonstration consistency*) and learner guidance is applied to stress the importance of alternative points of view and using multiple forms of media (*learner guidance*), that should not compete with each other for the learner attention (*relevant media*);
4. *Application Phase (let me)*: learning is promoted when learners are required to practice, consistently with the intended goals of the instruction (*practice consistency*), their new knowledge to solve a variety of problems (*varied problems*), with appropriate coaching (*diminishing coaching*). A single problem is not sufficient for learning a cognitive skill;
5. *Integration Phase* learning is promoted when new knowledge is integrated into the learner’s world, beyond the instructional environment: provide an opportunity for learners to publicly demonstrate their newly acquired skills (*watch-me*). For example, computer games do it increasing skill level that is apparent to the player. Learners need the

⁵ www.robocupathomeedu.org/challenges/2019-eu-homeedu.

Fig. 2 Real-world problems. From top left: inertia, relative motions (same and different direction), friction, circular motion

opportunity to defend and reflect on their new knowledge (*reflection*), and to modify it for use in their everyday lives (*creation*).

The application of these principles to our activities is briefly reported in the following, where we consider as *new knowledge* the activity of *programming a mobile robot to perform physics experiments*.

4.1 Problem-Centered Scenarios

The educational activities developed within our project allowed learners to be engaged in solving real-world problems involving physics concepts. Some examples of experiments arising from real-world observations are provided below and illustrated in Fig. 2, where MARRtino robot and other tools (including another mobile robot) are used to demonstrate different effects of Newtonian physics, such as inertia, friction, relative motions, etc.

Experiment n1. “*If I put a ball on the robot, does it move when the robot moves forward?*”. Using the robot and balls of different materials, dimensions and weight, students can program the robot to perform experiments on inertia, aimed at demonstrating (by direct visualization) that the ball, initially at rest and in the absence of external forces, maintains its initial position, seen by the reference system of the observer (the Earth), even during the robot movement.

Experiment n2. “*How the motion of two moving robots is composed in a given reference frame?*” The motion of an autonomous moving object (like a battery-powered toy vehicle with constant low speed) on top of the mobile robot moving in different directions allows to visualize such movements’ composition and observe motions in different

Fig. 3 Demonstration phase moment

reference frames. In this case, the students perform experiments on vector composition and on the relativity of the reference frames.

Other examples of Newtonian physics experiments are available on the project website.⁶

4.2 Activation of Existing Knowledge.

The students participating in the Lab2GO-Robotica project came from several Italian high schools where they have already attended a physics course, the past two years (66 hours each year). The physics topics they studied were, among others: the vectors and the composition of movements, the three principles of dynamics, the various types of movement (uniform, accelerated, various), the types of forces. The project aims to promote students' acquisition of new knowledge while building onto the existing knowledge about physics topics (i.e., programming a robot to perform physics experiments).

Conversely, students involved in the project had no experience with robotics and programming languages. Therefore, programming a mobile robot is considered as new knowledge.

4.3 Demonstration of New Knowledge.

The students were challenged with a question, for example: "*What happens when I put a ball on a robot and the robot starts moving forward?*". Then, the expert tutors proposed

⁶ <https://sites.google.com/a/dis.uniroma1.it/asl-robot/attivita/attivita1819>.

Fig. 4 Application phase moments

a video with the real robot to show the situation. Such demonstrations or showcasing are very important to give a clear and final objective to the teams. During this phase, the robot acted as a mindtool, as defined by Jonassen et al. (1998), because the students were actively engaged in interpreting the external world and reflecting on their interpretations (see Fig. 3).

4.4 Application of New Knowledge

The students' experiments were performed by programming the robot MARRtino to reproduce the real-world problem proposed by the expert tutors' demonstrations (see Fig. 4). During this phase, the students were the designers of the experiment with the robot, which once again acted as a mindtool, as defined by Jonassen et al. (1998), engaging learners in critical thinking and problem solving.

The expert tutors walked around the team areas to listen to comments, and provide hints if required (especially in programming the robot), but without giving explicit information about the challenge, thus appearing neutral to all the points of view of the considered problem (the *coaching approach* suggested by Merrill (2002)). They also encouraged the learner-control of the robot, not the teacher-control of the robot because, according to Jonassen (1996), mindtools fail when used for "traditional academic tasks set by teachers".

4.5 Integration of New Knowledge

In our project the integration phase was the time during which the teams worked independently at school or at home (for a total of 18h) to create original physics experiments

Fig. 5 Meeting structure

and apply the new knowledge on the mobile robot (*reflection, creation*). They received remote assistance from expert tutors and wrote the documentation about their project. At the same time, the teams were invited to visit robotics events for an amount of 12 hours to increase their motivation and engagement to the project (for example the RomeCup robotics Competition or the Maker Faire event). Lastly, the teams' experiments were publicly shown during a final exhibition event at the University and documented on websites, e.g. the Lab2GO-Robotica⁷ or the school website (*watch-me*). The activities done during the integration phase were not the focus of our assessment but, such demonstrations of student projects are very important to give a clear and final objective to the teams and motivate them to produce a tangible output of their activities as reported by Merrill's review of instructional design theories Merrill (2002).

In summary, our instructional method implemented the first principles of instruction in several educational activities, each one targeting a specific concept related to Newtonian physics. A typical flow of such activities carried out during each meeting is summarized below.

- During the meeting, expert tutors explained the problem, the experimental setup, and ask a question to the students. Possibly, some videos showing the experience were shown to the students as well. For example:

⁷ <https://web.infn.it/lab2go/informatica-e-robotica/>.

- Problem: *“How it changes the final displacement of a moving object on a moving platform, as the reference system changes?”*;
- Experimental setup: *“Place a squared paper on top of the MARRtino robot, and then put a mobile toy-robot on top of it. Make the two robots move in the same direction.”*;
- Question: *“Measure the displacement of the toy-robot concerning your reference system and the reference system of the toy-robot.”*.
- Each team of students set up the experiment and programmed the MARRtino robot to make simple movements (for example, move forward, turn, move in a square, etc.).
- The student teams turned on the mobile toy-robot and executed the program to move MARRtino robot.
- They measured the displacement of the mobile toy-robot at the end of the MARRtino motion (see Fig. 5).
- After one hour, all the teams were called to collectively discuss their solutions, answers and motivations.
- Finally, the tutors summarized the final answer to the question, clarifying all the involved aspects (e.g., using different reference systems in this case).

5 Instructional Method - Execution

The method defined above has been implemented during the 2018/2019 school year under the Lab2GO-Robotica project, a discipline within the more general Lab2GO program⁸ Organtini et al. (2017), that is a hands-on course for high school students. This program is part of the school-work alternation mechanism (*Alternanza scuola-lavoro (ASL)*), promoted in Italy by the Ministry of Education since 2015. The ASL mechanism allows the students to participate in project activities proposed by companies, public institutes, universities, etc. Lab2GO-Robotica was one of the options for the students and many schools requested to participate in this program.

The use of the ASL mechanism has both advantages and disadvantages.

Being an official program, students involved in these initiatives are rewarded with school credits that are necessary to complete their course of study. Agreements between project providers and schools are clearly defined, covering for example all insurance and responsibility aspects. Proposed projects are selected and students, teachers and students' parents can trust the hosting institutions.

For our research purposes, the drawback mainly relies on the fact that the students have to choose the program they want to attend, thus constraining the recruitment phase of experimental educational activities. For instance, researchers cannot decide which students can attend a specific project.

An alternative way of recruiting students outside this ASL program would have introduced considerable difficulties; most importantly, the students' possible lack of commitment for activities not included in their study plan.

Consequently, in this project, we decided to use the ASL mechanism, accepting the limitation of self-selection (by students that chose the program for themselves) or administrator selection (e.g., teachers) or both of these routes. In this way, we did quasi-experimental

⁸ <https://web.infn.it/lab2go/>.

Table 2 Parameters of the project

Who	61 students from 7 high schools divided by teams of 4
Why	Develop skills in physics, technology, Attitudes
Where	University, school, home, robotics events
robot role	tool to engage students to learn
Teacher role	supervisor, learner
Univ. Professor role	expert tutor
Background knowledge	not required (both for students and teachers)
robot cost	low cost product, affordable for public schools

Fig. 6 Organization of Lab2GO-Robotica project

research, according to the meaning of White and Sabarwal (2014), because of the lack of random assignment of the students to the experimental group. We identified a comparison group formed by the classmates of the experimental group students. In the first study, reported in Sect. 6, we analyzed if the comparison group was as similar as possible to the experimental group in terms of baseline characteristics (pre-intervention). In Sect. 6 we also analyzed the impact of this organization on the overall assessment of our project's results.

In this way, however, we believe that the results obtained in our study reflect the real world settings of the education system (at least in Italy) and can be easily replicated in a similar setting.⁹

During the execution of the project, the school teachers acted as supervisors. Although they were welcome to follow the project activities, they did not have an active role in the educational activities, since according to Jonassen (1996) cognitive tools should be “learner-controlled not teacher-controlled” because they fail when used for “traditional academic tasks set by teachers”.

The course started in November 2018 gathering 7 different public high schools in Italy, with a total of 61 students. Table 2 summarizes the project's characteristic in terms of the number of participants, robot and teacher roles, domain and location of the learning activities, that are fundamental parameters to describe an experiment, as underlined by Lindh et al. and Mubin et al. Lindh and Holgersson (2007), Mubin et al. (2013).

⁹ Indeed the program continued in the school year 2019/20, but was interrupted by COVID-19 emergency.

Fig. 7 Phase 1: students assemble the components to build MARRtino educationalRobot

5.1 Experimental Protocol

The project was organized in 4 Phases (0, 1, 2, 3), where Phase 2 contained the educational activities of programming robots for performing physics experiments that we wanted to assess in our study.

In contrast, the other Phases were provided to complete the project.

Students participating in the project were subdivided in an experimental and a comparison group, as described in Fig. 6.

The experimental group contained 29 students who participated in the activities of programming the robot for performing physics experiments (Phase 2) and preliminary activities on building the robot (Phase 1).

The comparison group contained 32 students who did not participate in the activities of programming the robot for performing physics experiments. This comparison group was composed of two sub-groups: 17 students (classmates of the Phase 1 students) and 15 students participating in Phase 1 (building the robot and learning basic robot programming, but without executing physics experiments).

In parallel, all the students (both groups) attended curricular physics lessons at school with their teachers (57 hours from November to March), who were not related to our project, but covered Newton's laws. The first study, reported in Sect. 6, analyzes the impact of this organization on the overall assessment of our project's results.

To validate the improvement in physics conceptual learning, we choose the FCI survey Hestenes et al. (1992) because it assesses Newton's first law, the superposition principle and the Newton's second law, that can be easily explored using a mobile robot like MARRtino.

We administered the Italian version of the FCI questionnaire to all the students twice: at the beginning (Phase 0) and the end (Phase 3). Pre-test and post-test data were then thoroughly analyzed to observe if and how the project changed some of the misconceptions about physics concepts, highlighted by the pre-test, and by comparing the results of the experimental and comparison groups. The results of these studies are reported in Sects. 7 and 8.

Detailed descriptions of the activities carried out in each of the Phases are given in the next sections.

5.2 Phase 0

The goal of Phase 0 was to administer the pre-test questionnaire to assess knowledge about physics concepts, technology attitudes, and collaboration work ability. In this phase, both experimental and comparison group students, a total of 61 students, attended short seminars about the state of the art of robotics and artificial intelligence, the project organization, and filled the pre-test questionnaire.

5.3 Phase 1

The goal of Phase 1 was to build the mobile robot and teach basic notions of programming, which are the prerequisites to attend Phase 2, during which the students programmed the robot to solve physics problems. Phase 1 was attended by 44 students. It involved one or two teams of students of the same classroom, available to participate in weekly meetings at the University or at their school. Each team was composed of 4 students and had a MARR-tino robot available. Phase 1 activities were done at the University, in 4 weekly meetings of 4 hours, for a total of 16 hours, where the students were supported by University professors and two undergraduate robotic master students. During the meetings, the students built the robot and installed the software required to control it (4 hours). They learned the basics of Python programming and how to run a simulator (12 hours). Some images showing such activities are depicted in Fig. 7. The teaching material given to the students in this phase is collected in the project website.¹⁰

After Phase 1, only 29 students attended Phase 2, joining the experimental group. This was due to reasons not related to our project and outside our control. Anyway, we could keep the remaining 15 students involved in the overall project, moving them to the comparison group.

It is important to observe that from the experimental results we did not notice significant differences between the two sub-groups of the comparison group, i.e., the 15 students coming from Phase 1 and the 17 classmates. Therefore, we believe that the comparison group was suitably composed to evaluate the performance of the experimental activities described in this article. Notice that these 15 students did not attend any lecture or perform any experiment related to physics, so they can be correctly considered part of the comparison group, since they did not participate in the educational activities that we want to measure.

5.4 Phase 2

The goal of Phase 2 was to program the robot to solve physics problems. The activities in Phase 2 are what we want to assess and thus have been performed only by the experimental group composed of 29 students. This phase lasted 16 hours (4 meetings of 4 hours each) and could be considered a reinforcement of school physics lessons using a new experimental methodology based on executing physics experiment through a mobile robot programmed by the students.

During this Phase, several experimental activities were carried out by the students, as described in the previous section, while the expert tutors submitted challenging questions

¹⁰ <https://sites.google.com/a/dis.uniroma1.it/asl-robot>.

to the learners and then let them build, program, and experience the solution with the mobile robot. During each meeting, both *demonstration* and *application* phases were implemented.

The demonstration phase lasted about 10 minutes and was open to questions by the students. The application phase lasted about one hour, including breaks when needed (see Fig. 4).

After the application phase, a 10 minutes open discussion was dedicated to asking feedback about the experiments with the robot (what happened? why?). Several examples of such experiments are collected in the program website.¹¹

Similarly to Hake, that computed the parameter f , as the fraction of class time spent on mechanics tm over the total semester time ts Hake (1998):

$$f = tm/ts$$

we computed the fraction of student's time spent on Phase 2 to the physics lesson time spent from November to March at school (equivalent to 57 hours):

$$f = 16/57 = 0.28$$

We will use this value for further considerations later on.

5.5 Phase 3

The goal of Phase 3 was to administer the post-test questionnaire to assess knowledge about physics concepts, technology attitudes, and collaboration work ability. Both experimental and comparison group students, a total of 61 students, participated in this phase.

6 Study I - Pre-Test Analysis

In this section, we present the first analysis of data taken from the pre-test questionnaires collected during Phase 0. The goal of this first study is to analyze some characteristics of the sample of students.

6.1 Procedure

At the end of Phase 0, we described the experimental protocol to the students, explaining the use of questionnaires before (pre) and after (post) the project activities. No personal information was collected, the questionnaires were fully anonymous, using a nickname to match pre-test and post-test data to measure learning improvements. The students were allowed to withdraw from the study at any time without providing a reason. The pre-test questionnaire included mainly closed-ended questions related to student's technology attitudes, their perception of collaborative work attitudes, and their knowledge about physics concepts (i.e., the FCI test). There were also two open questions to know the student's expectations about the project contents and learning.

¹¹ <https://sites.google.com/a/dis.uniroma1.it/asl-robot>.

Table 3 Analysis of pre-test scores comparing experimental and comparison groups. In bold the significant statistical p -values < 0.05

Dimension	Experimental		Comparison		p -value
	Mean	Std.Dev.	Mean	Std.Dev.	
Knowledge about physics concepts	27.2 %	11.2 %	20.7 %	8.0 %	0.017
Attitude towards technology sector	4.154	0.518	3.731	0.805	0.025
Willingness to improve their work	4.000	0.462	3.613	0.789	0.032
Perception of technology	4.173	0.509	4.145	0.608	0.853
Attitude towards the University	3.154	0.527	3.108	0.467	0.726
Self-Efficacy in technology sector respect to Gender	2.731	0.587	2.984	0.713	0.154
Collaboration	3.831	0.418	3.832	0.552	0.991
Use of technology	3.856	0.516	4.000	0.487	0.283
Reflexivity	3.744	0.560	3.753	0.558	0.951
Autonomy	4.019	0.490	3.774	0.534	0.079

6.2 Sample

We administered the pre-test questionnaires to 61 students (age $M = 16.18$, $SD = 0.73$), including 9 females.

6.3 Data Analysis

The following domains were assessed by using three types of questionnaires: (1) *knowledge about physics concepts*, (2) *technology perceptions and attitudes*, (3) *students' perception of collaborative work attitudes*.

Knowledge about physics concepts were assessed in pre-test and post-test using the Force Concept Inventory v95 (FCI), a questionnaire with 30 five-alternative multiple choice questions (only one is correct), available on the *PhysPort* website.¹² The *PhysPort* is developed by the American Association of physics Teachers (AAPT), in collaboration with Kansas State University and supported by the National Science Foundation.¹³ We used the Italian version of the survey, provided by the *PhysPort* website.

The FCI questionnaire assesses students' understanding of the most basic concepts in Newtonian physics, using everyday language; it is characterized by having, for each question, one correct answer and four alternatives. The wrong answers are built upon *distractors* that are concepts based on an initial common-sense knowledge of students, i.e., a system of beliefs and intuitions about physical phenomena derived from personal experience. From previous research, we know that the initial common sense knowledge of students has a significant effect on their performance in physics courses, and conventional instruction is ineffective in correcting defects in this knowledge (nevertheless the best physics teacher) Halloun and Hestenes (1985). The FCI is appropriate for Intro college and High school students. This test includes clusters of questions by topic that help to understand better for

¹² <https://www.physport.org/assessments/assessment.cfm?I=5&A=FCI>.

¹³ <https://www.physport.org/webdocs/About.cfm>.

Table 4 FCI questions related to the first law, second law and superposition principle

FCI Dimension	Question
First law	6, 7, 8, 10, 11, 17, 23, 24, 25
Second law	8, 9, 21, 22, 26
Superposition principle	8, 9, 11, 17, 25

which conceptual dimension the students need more help. The topics are: Newton's first, second, and third laws, kinematics, superposition principle, and kind of forces. FCI survey was chosen because it assesses the first law, the superposition principle, and the second law, concepts that can be easily explored using a mobile robot like MARRtino. FCI scoring was done following the PhysPort Expert Recommendation on Best Practices for Administering Concept Inventories.¹⁴

Technology perceptions and attitudes were assessed in the pre-test with a questionnaire composed of 11 items, re-adapting the one used in Sowell et al. (2016). After Factorial analysis, the items were grouped in 4 scales: "Attitude towards technology", "Attitude towards the University", "Perception of the technology sector" and "Self-Efficacy in technology sector respect to Gender". For each item, an agreement/disagreement scale with five-level Likert scale was provided.

Students' perception of collaborative work attitudes were collected in the pre-test with a questionnaire composed of 20 items, re-adapted from the CKP questionnaire (Collaborative Knowledge Practices Muukkonen et al. 2017), that is inspired by the Trialogical Learning Approach Paavola et al. (2004), Paavola et al. (2006), Sami and Kai (2009). After Factorial analysis, the 20 items were grouped in 5 scales: "Collaboration", "Use of technology", "Willingness to improve their work", "Autonomy", "Reflexivity". Also in this questionnaire, for each item, an agreement/disagreement scale with five-level Likert modes was provided.

6.4 Results of Pre-Tests

We analyzed the pre-test data using a two-tailed T-test analysis and reported the results in Table 3. Statistical difference between the experimental group and the comparison group (i.e., p -value < 0.05) has been found in three dimensions: *Knowledge about physics concepts* (i.e., FCI questionnaire), *Attitude towards the technology sector*, and *Willingness to improve their knowledge*, while for the other dimensions no significant difference have been registered.

The difference can be explained with the selection procedure that was imposed by the choice of the ASL mechanism, as already mentioned in the previous section. This difference must be considered when drawing the conclusion of our study, since students with slightly better physics knowledge, attitudes towards technology and motivation were involved in the experimental group. Nevertheless, it is important to observe that current practices introducing educational robotics activities in a real context always brings some bias (in some cases not even measured), as it is not possible to work with a large sample of students properly selected

¹⁴ www.physport.org/expert/AdministeringConceptInventories/.

Fig. 8 Student's scores of the experimental group (above) and comparison group (below) to the 30 FCI questions. The dashed curve represents the pre-test data, the solid one represents the post-test data

and fully motivated to participate in such experiments. Moreover, as shown in Sect. 7, even by considering only the experimental group, there are significant differences and large effect size between pre-test and post-test scores, thus showing an actual benefit of the proposed instructional method for this group of students.

Fig. 9 Experimental group: percentage of the correct answers (Score) to all the 30 FCI questions, for pre-test and post-tests

7 Study II - Physics Score Analysis and Course Effectiveness

As this article focuses on teaching the first and second laws and the superposition principle using a mobile robot, we deeply analyzed the score in these dimensions. The specific questions related to these dimensions are described in the FCI Implementation Guide¹⁵ and are reported for convenience in Table 4.

7.1 Procedure

At the end of the project (March 2019), during Phase 3, all the 61 students were invited to the University to fill the post questionnaires, including the FCI survey. Moreover, to allow students to give feedback on the course, we included two open-ended questions, to know if their expectations were met.

7.2 Sample

The pre-test and post-test questionnaires completed and returned were matched using the nickname. The response rate of the experimental group was 90% because 26 pre-test and post-test questionnaires were completed and returned.

7.3 Data Analysis

Students' scores have been analyzed with the FCI scoring tool (Assess SpreadSheet v7a) and the PhysPort Assessment Data Explorer¹⁶ to assess the learning gains obtained by participating to our project.

¹⁵ <https://www.physport.org/assessments/assessment.cfm?I=5&A=FCI>.

¹⁶ www.physport.org/explore/FCI.

Fig. 10 Experimental group: pre-test and post-test correct answers percentage for all the 30 FCI questions, clustered by conceptual dimensions. Error bars are one standard error of the mean

Fig. 11 Experimental group: correct answers percentage for first law dimension (9 questions)

7.4 Score Analysis

The analysis of the scores between pre-test and post-test are shown in a graphical form in Fig. 8.

The experimental group's average score of correct answers, in all the FCI dimensions, shows a significant difference, increasing from 27.2% (Stdev 11.2%) to 36.4% (Stdev 14.2 %) (p -value = 0.013). While the scores in the comparison group show less significant difference: from 20.7% (Stdev 8.0%) to 23.2% (Stdev 6.7%) (p -value = 0.21).

These results confirm a significant improvement in FCI score due to the activity of programming a robot for performing physics experiments, in comparison with traditional curricular physics lessons at school.

In the rest of this section, we examine the results of the experimental group in more detail.

Table 5 FCI normalized change values, considering all the FCI dimensions, as well as the three dimensions explored with MARRtino robot

	$\bar{c}$
All the 6 FCI dimensions	0.07
First law	0.18
Second Law	0.10
Superposition Principle	0.20

Table 6 Effect size values, considering all the FCI dimensions, as well as the three dimensions explored with MARRtino robot

	Effect Size
All the 6 FCI dimensions	0.69
First law	0.70
Second Law	0.61
Superposition Principle	0.80

7.4.1 Experimental Group Results

The histogram in Fig. 9 represents the percentage of the correct answers (Score) of the pre-test and post-test to all the 30 FCI questions for the experimental group. In the graph, the zones of the dominance of Newtonian physics, as interpreted by Hestenes and Halloun (1995), are also displayed. In the pre-test, the majority of students obtained scores in the range 10–29%, while in the post-test the majority of students obtained scores in the range 30–49%, showing an improvement of the universal force concept and the identification of active and passive agents of force. Moreover, some students got scores in the range $\geq 60\%$, which represents the entry threshold of the Newtonian universe, where coherent dynamical concepts, like velocity vectors, acceleration, and force, are developed.

Figure 10 shows pre-test and post-test percentage of FCI correct answers, clustered by conceptual dimensions. Notice that first and second law and the superposition principle dimensions are the ones with the bigger improvement in the post-test, as expected because the challenging physics tasks, discussed during the meetings, were about these three dimensions.

Figure 11 shows detailed results for the first law dimension. It is interesting to notice that, in the post-test, more than half (i.e., 56%) of the students got a percentage of correct answers $\geq 60\%$, that corresponds to moderate dominance of the Newtonian universe zone.

7.4.2 Effectiveness of the Course

To measure the overall effectiveness of our educational robotics activity, we considered the students who took both the pre-test and the post-test and computed the following performance measures: the average *normalized change*; the *effect size*; the distribution of the *raw gain* versus the pre-test score. Then we compared our results with other specific initiatives in education reported by the physics education research community.

The average normalized change is computed as the average of the normalized change for all the students in the experimental group. The normalized change, c , is defined by Marx and Cummings (2007) as follows:

Fig. 12 Distribution of raw gain ($\%Post - \%Pre$) vs. pre-test. Comparison between the results of our experiment (solid circles), and other specific initiatives in physics education reported by Hake: interactive engagement courses (IE), and traditional courses (T)

$$c = \begin{cases} \frac{\%Post - \%Pre}{100 - \%Pre} & \%Post > \%Pre \\ 0 & \%Post = \%Pre \\ drop & \%Post \text{ and } \%Pre = 0 \text{ or } 100 \\ \frac{\%Post - \%Pre}{\%Pre} & \%Post < \%Pre \end{cases}$$

The average of the normalized changes, $\bar{c}$, is then computed as the average of the c values of all the students. The results of this analysis are summarized in Table 5, considering all the FCI dimensions, as well as the three dimensions explored within our project: all the $\bar{c}$ values are in line with the ones obtained by Hake (1998), Von Korff et al. (2016) that used the FCI test to assess traditional teaching methods.

To better assess how substantially the student's knowledge of Newtonian physics has changed as a result of the project, we also computed the effect size of the difference between the pre-test and post-test. Indeed Nissen et al. (2018) recommends using the effect size rather than the normalized gain or normalized change for measuring student learning, to avoid the bias towards high pre-test scores.

An effect size is a measure of how important a difference is: large effect sizes mean the difference is important; small effect sizes mean the difference is unimportant. It normalizes the average raw gain in a population, defined as $(\%Post - \%Pre)$, by the standard deviation in individuals' raw scores, giving a measure of how substantially the pre-test and post-test scores differ.

We calculate the effect size using the Cohen's formula:

Fig. 13 The S-C plot shows the score and concentration results of individual multiple choice questions

Table 7 The three-levels coding scheme for classification of student model states, using the score *S* and the concentration factor *C*. L = Low, M = Medium, H = High

<i>S</i>	Level	<i>C</i>	Level
0–0.4	L	0–0.2	L
0.4–0.7	M	0.2–0.5	M
0.7–1.0	H	0.5–1.0	H

$$d = (\%Post - \%Pre) / Stdev_p$$

where $Stdev_p$ is the pooled standard deviation of the raw gain as a percent:

$$Stdev_p = \sqrt{(StdevPre^2 + StdevPost^2) / 2}$$

A suggested scale for the effect size is small (0.2), medium (0.5), and large (0.8).

As shown in Table 6, considering all the FCI dimensions, as well as the three dimensions explored within our project, the effect size of our project is between medium and large.

Finally, the distribution of the raw gain ($\%Post - \%Pre$) versus the pre-test score, for the experimental group, is reported in Fig. 12, in comparison with other specific initiatives in physics education reported in Hake (1998). The big solid circle represents the average on all the FCI dimensions. The small grey, yellow and green solid circles represent the average respectively of the superposition principle, of the second law, and the first law. They reached similar raw gain values of the traditional courses (triangles), indicating that by programming a mobile robot it is possible to learn physics concepts related to the first and second laws and the superposition principle with a learning gain comparable with the one obtained with traditional lessons.

It is important to also notice that, in comparison with other approaches, the course time of our instructional method was significantly shorter. Indeed, when referring to the fraction of time spent in the instructional method, for our project we have a value

Fig. 14 Experimental group: average of all experimental group student's responses to the 30 FCI questions, given respectively in the pre (circle) and in the post-test (triangle)

$f = 0.28$ (see Sect. 5.4), whereas for the studies reported by Hake f ranged from 0.7 to 1.0 Hake (1998). Therefore, we can state that, in comparison with other approaches, our project achieved at least the same results as the traditional methods, but in a much smaller time frame.

The quantitative analysis of the results in physics learning score and course effectiveness reported in this section can be summarized as follows: (1) our project provided for a significant learning gain in physics, (2) more than half of the students achieved a score on the first law above the threshold of moderate dominance of Newtonian physics, (3) effect size of the project has been medium/large, 4) in comparison with other approaches, the results of our project (especially in the three dimensions analyzed) are at least as good as traditional methods.

8 Study III - Physics Score Concentration Data Analysis

In our third study, we analyzed how our project changed the distribution of the answers to the FCI questions, following the Pea idea Pea (1985) that educational and cognitive tools (MARRtino robot in our case) can reorganize the mental schemata of learners.

To this end, we performed a *concentration analysis* of data, that gives information, for each question j , on how the answers are distributed among the different available choices i Bao and Redish (2001).

In our project we used the FCI test that has 30 five-alternative multiple choice questions, so j takes integers in $[1, 30]$ range, i takes integers in $[1, 5]$ range, and the concentration factor, C_j , is calculated as follows:

$$C_j = \frac{\sqrt{5}}{\sqrt{5} - 1} \left(\frac{\sqrt{\sum_{i=1}^5 n_i^2}}{N} - \frac{1}{\sqrt{5}} \right)$$

where N is the total number of students and n_i is the total number of students who selected choice i .

Fig. 15 Experimental group: distribution of the S-C data of pre-FCI answers (circles) and post-FCI answers (triangles). The mean values are also shown

Fig. 16 Experimental group, first law: distribution of pre-FCI answers (circles) and post-FCI answers (triangles) to the first law questions in the S-C plot. The mean values are shown by the solid circle and triangle

C_j takes values in $[0,1]$ range, with $C_j = 0$ corresponding to an equal distribution of answers among all the possible choices (random guess) and $C_j = 1$ corresponding to all students selecting the same answer i .

For each question, by combining the concentration factor C_j with the score S_j (the number of students who selected the correct answer), it is possible to analyze the response patterns of a group of students and to provide a measure of students' performance by indicating whether the question triggers a common misconception Bao and Redish (2001), Cavik and Kurnaz (2019). Moreover, the pattern of the shift from pre-instruction to

Table 8 The three-levels coding scheme for classification of student model states, using score S , and the concentration deviation Γ . L = Low, M = Medium, H = High

S	Level	Γ	Level
0–0.4	L	0–0.2	L
0.4–0.7	M	0.2–0.4	M
0.7–1.0	H	0.4–1.0	H

Fig. 17 S - Γ pre and post-test data for all the FCI questions: a comparison between our experiment (EXP) and other specific initiatives in physics education (BAO-T, BAO-IE). IE = interactive engagement courses, T = traditional courses

post-instruction tells how the “state” of the group of students evolves with the instruction Bao and Redish (2001). Due to the constraint between the score and concentration factor, data points can only exist in the area between the two boundary lines showed in Fig. 13.

Each combination of the pairs of S-C levels enables a set of possible student’s answers states, represented by one, two, or non-peak regions, depending on the C level reached up. The three-levels coding scheme for the classification of student model states is indicated in Table 7 and Fig. 13.

For example, the “one-peak” situation is typical for either an LH or an HH type of response. In fact, in the LH case, students have low scores and most of them picked the same distractor. Therefore, it could be considered as a strong indication that the question triggers a common incorrect model. In the HH case, students have high scores and most of them picked the same answer which is correct.

The “two-peaks” situation happens when many of the responses are concentrated on only two choices. If one of the two is the correct answer, the response type is MM; if both choices are incorrect, the response type is instead LM.

The “no-peak” situation happens when student responses are somewhat evenly distributed over three or more of the choices. The response pattern is usually LL. This implies that most of the students do not have a strong preference for any model on the topic and the responses are close to random guesses.

Fig. 18 Experimental group, first law: S- Γ average values of all the questions, for pre and post-test

Fig. 19 Experimental group, first law: S- Γ average value for each question

We performed the physics score concentration data analysis for the experimental group students. The two points in Fig. 14 represent the average of all experimental group student's responses to the 30 FCI questions, given respectively in the pre and post-test. The average of post-test answers generally improves in the S-C plot, in terms of higher score. It is possible to appreciate the distribution of the answers in Fig. 15, where each point on the graph represents the average of all student's responses to each pre (circles) and post (triangles) FCI question, while the shift vector on the graph just want to help the reader to look for the average post-test value (big triangle) corresponding to the pre-test value (big circle). Indeed the shift towards the direction of a higher score indicates that more students chose the correct answer, moving towards the MM region.

The analysis focused on the first law topic is shown in Fig. 16 that shows the distribution of pre-FCI answers (circles) and post-FCI answers (triangles) to the first law questions in the S-C plot. Each point in the graph represents the average of all student's responses to one FCI question. The question's number is reported near the point and the shift from pre-test answer to post-test answer is evidenced by an arrow.

The mean values are represented by the solid circle and triangle. The average of pre-test is in the two-peak LM region, but the post-average moves from the LM region to the MM one, confirming the learning improvement on this topic. There is a score improvement

Fig. 20 First law, question 8: the scattering of the answers, from pre to post-test. Choice B is the right answer to the question

for all the questions, outlined by arrows pointing to the right, and a greater concentration (upward pointing arrows), in some cases even toward the HH region (questions 7, 8 and 10), indicating that more students chose the correct answer.

8.1 Concentration Deviation

We can learn more about students' wrong answers by also looking at the S – I distribution. I is called *concentration deviation*, and it is not affected by the score S , thus giving the concentration for student's incorrect answers Bao and Redish (2001). For each question j , the concentration deviation, I_j , is calculated as follows:

$$I_j = \frac{\sqrt{4}}{\sqrt{4} - 1} \left(\frac{\sqrt{\sum_{i=1}^5 n_i^2 - n_s^2}}{N - n_s} - \frac{1}{\sqrt{4}} \right)$$

where N is the total number of students, n_i is the number of students who selected choice i , and n_s is the number of students who selected the correct answer. I_j takes values in the full range $[0, 1]$, without restriction area.

The three-levels I coding scheme for the classification of student model states is indicated in Table 8. High I values (> 0.4) indicate a distractor, i.e., situations where student's responses have alternative models. Low S with high I values means that the question triggers a common incorrect model and an effective distractor (LH region); high S with high I values means that the distractor is not much effective (HH region).

Figure 17 compares the S – I values for the experimental group in our project (EXP) with other specific initiatives in physics education (BAO-T, BAO-IE) reported in Bao and Redish (2001). The experimental group reached a post-test I value (solid triangle) that is almost equal to the values obtained by Bao, that used methods other than robotics (traditional lecture-based lessons (T) and interactive engagements lessons (IE), as defined by Hake (1998)).

For the first law, Fig. 18 shows that the post-test average value moves from the LM region toward the MM one, for the experimental group.

We can see more details in Fig. 19, where each point in the graph represents the average of all student's responses to a first law question, while the shift vector on the graph just want to help the reader to look for the post-test value corresponding to a pre-test value.

Fig. 21 First law, questions 6, 8, and 17: the black sector represents the correct answer percentage. Notice how the concentration on the wrong answers (white sectors) changes over time, from pre-test to post-test: for example, some options are reduced or disappear in the post-test

In the pre-test, 56% of answers are in the low score range (L) and the low and high I range (L, H), producing an average value in the LM region (the big circle). The post-test data move towards higher S and I values, generating an average value in the MM region (the big triangle). It means that the students' answers are concentrating only on a few distractors. In some cases, post S - I data are in the HM region (questions 7 and 10) or entirely inside the HH region (question 6), meaning that the students' answers concentrate on two or one weak distractor. For example, option A represents a weak distractor for question 6 (see the post-test results of item 6 in Fig. 21).

When the students' post-test incorrect responses are all significantly reduced, then the I value remains constant while the score S increases, respect to the pre-test data (see questions 8 and 24). Notice how question 23 moves away from the LL region (random answers, no mental model) toward the MM region (two prevalent mental models); questions 11 and 25 moves from high to medium I values increasing the score, meaning that the students changed their mental model towards the correct one.

Further details are illustrated in Figs. 20, 21, that analyze some specific questions (6, 8, and 17). In these figures it is possible to observe the following outcome: some wrong answers are no longer chosen in the post-test, for example choices C and E for question 8 (Fig. 20); some wrong answers are chosen less, for example choice A for question 17 and choice D for question 6 (Fig. 21).

In summary, the concentration analysis presented in this section confirms that the proposed instructional method can change the students' mental model.

9 Discussion and Lessons Learned

The students' commitment to programming the robots to see the experiments in operation had a fundamental impact on the motivations and involvement in carrying out the project, as already noted by Pea (1985), Jonassen (1996), Jonassen et al. (1998); Jonassen and Rohrer-Murphy (1999); Jonassen (2000). We believe that the good results in the development of the project have been mainly obtained thanks to capturing the students' attention and keeping them engaged in solving challenging problems with the robot, that worked as partner in learning. On the contrary, we observed that during sessions in which videos of the robots were shown to the students, attention and engagement were reduced.

In addition to the results in the questionnaires, at the end of the project, we obtained very positive feedback from the students. Within the experimental group, 92% of students stated that the project met their expectations. Among these, 72% indicated that they were fully satisfied. Moreover, students perceived to have learned teamwork (69%) and to have acquired new knowledge as robot building (35%), robot programming (27%), physics competence (23%), technology competence (19%).

The first lesson learned from our experience is that students can take advantage of technology to learn curricular subjects, under two conditions: (1) the technology made accessible to the students is an easy and effective way, demonstrating that they can actually manage and take advantage of it; (2) the students' levels of attention and engagement are kept high by a direct personal and group involvement in the educational activities.

We also collected positive teacher's feedback: they witnessed great motivation and engagement of their students and understood the educational value of the activity. Our observation was also that teachers' presence was very important to drive the students' activities.

The second lesson learned is that teachers are very valuable in activities involving new technological tools, even if they do not directly help the students understand the technology. Instead teachers can contribute in motivating, organizing, driving and keeping the students engaged on a project.

The analysis of the pre-test and post-tests questionnaires showed its utility in evaluating and assessing our instructional methodology and the use of the FCI survey, a well-known instrument in physics education, allowed for comparison with other initiatives. The results obtained so far provide a clear picture of the users' final point in this experimentation.

Moreover, our choice to introduce our educational project within an existing initiative approved by the Ministry of Education allows the project to be replicated in the next years and in other schools with minimal organization effort.

The third lesson learned is that the evaluation of an educational activity should be the key focus of the entire project and thus it must be carefully designed before the execution of the program. On the other hand, it is also important to properly adapt the educational program and the corresponding evaluation activities to the students' actual needs to avoid the execution and evaluation of programs that could not be replicated or extended to other cases.

10 Conclusions

In this article, we presented and measured the learning gain and effectiveness of a new instructional methodology where Newtonian physics concepts are empowered through programming a mobile robot. The robot acted as a mindtool, as defined by Jonassen et al. (1998), engaging learners in critical thinking and problem solving.

We applied this method to an experimental group of 29 students from 7 different high schools in Italy that performed experiments by programming the robot. A comparison group of 32 classmates was recruited to compare its results with the experimental group.

We filled some of the gaps described in Benitti (2012), as we assessed our methodology using quantitative data and statistical analysis. Moreover, we addressed a curricular subject (physics), involving 15–19 years old students during lesson time at school and within a recognized curricular program (not an extracurricular activity). We also implemented best-practices evidenced by educational robotics researchers: one robot kit for a small team of students (teamwork accelerator); realistic but affordable tasks (move the robot forward/backward); short theory lessons; suitable spaces for robot experiments (school and University laboratories).

The overall outcome, thoroughly analyzed from different perspectives (sample analysis, score analysis, concentration analysis, concentration deviation) and with different measures (normalized change, effect size, raw gain), confirms that mobile robot programming is a valuable experience to empower physics concepts. Students can effectively change their Newtonian mental model with an educational robotics activity that does not contain specific physics lessons. From these results, it is easy to foresee that a proper integration of educational robotics with specific lessons would increase learning performance even more.

All the results reported in this article clearly show the effectiveness of the use of educational robotics to empower Newtonian physics concepts and allow answering our original research question as follows:

Programming a mobile robot to perform physics experiments empowered the knowledge of Newtonian physics concepts, with performance at least as good as traditional lessons, but in a shorter time frame. The added values with respect to the traditional lessons are in terms of social behaviors (teamwork), new skills and competences acquisition (robotics, coding, physics, technology), and engagement with the 3D physical world.

We believe that this result will bring new ideas about using robotics to empower physics in high schools, for example defining new instructional methods in which robotics experiments are effectively integrated with specific physics lessons.

The instructional method, most of the technical developments and the evaluation methodology can be applied to other curricular subjects, to analyze effectiveness in programming mobile robots to gain knowledge in other subjects.

We also hope that our project's experimental methodology would inspire other similar studies to advance the research in educational robotics by allowing comparative analysis of different experiences.

Acknowledgements We would like to express our greatest appreciation to all the students who adhered to our project and to the teachers and the schools that supported them.

References

- Ahmed, I., Lubold, N., Walker, E. (2018). Robin: Using a programmable robot to provide feedback and encouragement on programming tasks. *Artificial intelligence in education* (pp. 9–13). Cham: Springer International Publishing.
- Alimisis, D. (2012). Robotics in education & education in robotics: Shifting focus from technology to pedagogy. In *Proceedings of the 3rd international conference on robotics in education* (pp. 7–14).
- Andre, T. (1997). Construction: Implications for instructional design. *Instructional Design: International Perspectives. Theory, research, and models. Vol. 1*, 1, 243.
- Bao, L., & Redish, E. F. (2001). Concentration analysis: A quantitative assessment of student states. *American Journal of Physics*, 69(S1), S45–S53.
- Benitti, F. B. V. (2012). Exploring the educational potential of robotics in schools: A systematic review. *Computers & Education*, 58(3), 978–988. <https://doi.org/10.1016/j.compedu.2011.10.006>.
- Cavik, E. E., & Kurnaz, M. A. (2019). Analysis of the responses of science teacher candidates to force concept inventory by concentration factor. *Universal Journal of Educational Research*, 7(1), 111–117.
- Church, W. J., Ford, T., Perova, N. & Rogers, C. (2010). Physics with robotics—using lego mindstorms in high school education. In *2010 aaai spring symposium series*.
- Ferrarelli, P., Lapucci, T. & Iocchi, L. (2017). Methodology and results on teaching maths using mobile robots. In Ollero, A., Sanfeliu, A., Montano, L., Lau, N. & Cardeira, C. B. (Eds.), *ROBOT 2017: Third Iberian Robotics Conference - Volume 2*, Seville, Spain, November 22–24, 2017 *ROBOT 2017: Third iberian robotics conference - volume 2, seville, spain, november 22-24, 2017* (Vol. 694, pp. 394–406). Springer. https://doi.org/10.1007/978-3-319-70836-2_33.
- Fraser, N. (2015, Oct). Ten things we've learned from blockly. In 2015 IEEE Blocks and Beyond Workshop (Blocks and Beyond) (p.49–50). <https://doi.org/10.1109/BLOCKS.2015.7369000>.
- Gardner, H. (2018). Multiple approaches to understanding. In *Contemporary theories of learning* (pp. 129–138). Routledge.
- Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American journal of Physics*, 66(1), 64–74.
- Halloun, I. A., & Hestenes, D. (1985). Common sense concepts about motion. *American journal of physics*, 53(11), 1056–1065.
- Hestenes, D., & Halloun, I. (1995). Interpreting the force concept inventory: A response to March 1995 critique by huffman and heller. *The Physics Teacher*, 33(8), 502–502. <https://doi.org/10.1119/1.2344278>.
- Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, 30(3), 141–158.
- Hussain, S., Lindh, J., & Shukur, G. (2006). The effect of lego training on pupils' school performance in mathematics, problem solving ability and attitude: Swedish data. *Journal of Educational Technology Society*, 9(3), 182–194.
- Jonassen, D. (2000). *Mindtools for schools: Engaging critical thinking with technology*. Columbus: Merrill.
- Jonassen, D. H. (1996). Learning with technology: Using computers as cognitive tools. *Handbook of research for educational communications and technology*.
- Jonassen, D. H., Carr, C., & Yueh, H.-P. (1998). Computers as mindtools for engaging learners in critical thinking. *TechTrends*, 43(2), 24–32.

- Jonassen, D. H., & Rohrer-Murphy, L. (1999). Activity theory as a framework for designing constructivist learning environments. *Educational technology research and development*, 47(1), 61–79.
- Lindh, J., & Holgersson, T. (2007). Does lego training stimulate pupils' ability to solve logical problems? *Computers & Education*, 49(4), 1097–1111. <https://doi.org/10.1016/j.compedu.2005.12.008>.
- Marx, J. D., & Cummings, K. (2007). Normalized change. *American Journal of Physics*, 75(1), 87–91.
- Merrill, M. D. (2002). First principles of instruction. *Educational technology research and development*, 50(3), 43–59.
- Mitnik, R., Nussbaum, M., & Soto, A. (2008). An autonomous educational mobile robot mediator. *Autonomous Robots*, 25(4), 367–382. <https://doi.org/10.1007/s10514-008-9101-z>.
- Mubin, O., Stevens, C. J., Shahid, S., Al Mahmud, A., & Dong, J.-J. (2013). A review of the applicability of robots in education. *Journal of Technology in Education and Learning*, 1, 209.
- Muukkonen, H., Ilomäki, L., Lakkala, M. & Toom, A. (2017). Adaptation of the collaborative knowledge practices questionnaire to uppersecondary education. In EDULEARN17 proceedings. <https://doi.org/10.21125/edulearn.2017.2185>.
- Nelson, L. M. (1999). Collaborative problem solving. *Instructional design theories and models: A new paradigm of instructional theory*, 2, 241–267.
- Nissen, J. M., Talbot, R. M., Thompson, A. N., & Van Dusen, B. (2018). Comparison of normalized gain and cohen's d for analyzing gains on concept inventories. *Physical Review Physics Education Research*, 14(1), 010115.
- Organtini, G., Ameli, F., Cavoto, G., Di Marco, E., Piacentini, F., Morganti, S. & Faccini, R. (2017). Promoting the physics laboratory with Lab2GO. In Edulearn17 proceedings (p. 5264–5268). IATED. <https://doi.org/10.21125/edulearn.2017.2185>.
- Paavola, S., Hakkarainen, K., & Sintonen, M. (2006). Abduction with dialogical and trialogical means. *Logic Journal of IGPL*, 14(2), 137–150.
- Paavola, S., Lipponen, L., & Hakkarainen, K. (2004). Models of innovative knowledge communities and three metaphors of learning. *Review of Educational Research*, 74(4), 557–576.
- Pasternak, E., Fenichel, R. & Marshall, A. N. (2017). Tips for creating a block language with blockly. In 2017 IEEE blocks and beyond workshop (b b) (p. 1–24). <https://doi.org/10.1109/BLOCKS.2017.8120404>.
- Pea, R. D. (1985). Beyond amplification: Using the computer to reorganize mental functioning. *Educational Psychologist*, 20(4), 167–182.
- Pfeifer, A., & Lugin, B. (2018). Female robots as role-models? - the influence of robot gender and learning materials on learning success. *Artificial intelligence in education* (pp. 276–280). Cham: Springer International Publishing.
- Salomon, G., Perkins, D. N., & Globerson, T. (1991). Partners in cognition: Extending human intelligence with intelligent technologies. *Educational Researcher*, 20(3), 2–9. <https://doi.org/10.3102/0013189X020003002>.
- Sami, P. & Kai, H. (2009). From meaning making to joint construction of knowledge practices and artefacts: A trialogical approach to cscl. In Proceedings of the 9th international conference on computer supported collaborative learning-volume 1 (pp. 83–92).
- Schank, R. C., Berman, T. R., & Macpherson, K. A. (1999). Learning by doing. *Instructional-design theories and models: A new paradigm of instructional theory*, 2(2), 161–181.
- Shiomi, M., Kanda, T., Howley, I., Hayashi, K., & Hagita, N. (2015). Can a social robot stimulate science curiosity in classrooms? *International Journal of Social Robotics*, 7(5), 641–652.
- Sowells, E., Waller, L., Ofori-Boadu, A. & Bullock, G. (2016). Using technology summer camp to stimulate the interest of female high school students in technology careers. In Management of engineering and technology (picmet), 2016 portland international conference on (pp. 1688–1696).
- Tan, J. T. C., Iocchi, L., Eguchi, A. & Okada, H. (2019). Bridging robotics education between high school and university: RoboCup@Home Education. In 2019 IEEE africon, accra, ghana, september 25–27, 2019 (pp. 1–4). IEEE. <https://doi.org/10.1109/AFRICON46755.2019.9133791>.
- Tanaka, F. & Matsuzoe, S. (2012). Children teach a care-receiving robot to promote their learning: Field experiments in a classroom for vocabulary learning. *Journal of Human-Robot Interaction* 1(1).
- Timms, M. J. (2016). Letting artificial intelligence in education out of the box: Educational cobots and smart classrooms. *International Journal of Artificial Intelligence in Education*, 26, 701–712.
- Tomara, M., Tselfes, V., & Gouscos, D. (2017). Instructional strategies to promote conceptual change about force and motion: A review of the literature. *Themes in Science and Technology Education*, 10(1), 1–16.
- Vaughan, R. (2008). Massively multi-robot simulation in stage. *Swarm Intelligence*, 2(2), 189–208.

- Von Korff, J., Archibeque, B., Gomez, K. A., Heckendorf, T., McKagan, S. B., Sayre, E. C., & Sorell, L. (2016). Secondary analysis of teaching methods in introductory physics: A 50 k-student study. *American Journal of Physics*, 84(12), 969–974.
- Vygotsky, L. S. (2012). *The collected works of ls vygotsky: Scientific legacy*. Berlin: Springer.
- White, H., & Sabarwal, S. (2014). Quasi-experimental design and methods. *Methodological briefs: Impact Evaluation*, 8, 1–16.
- Williams, D. C., Ma, Y., Prejean, L., Ford, M. J., & Lai, G. (2007). Acquisition of physics content knowledge and scientific inquiry skills in a robotics summer camp. *Journal of Research on Technology in Education*, 40(2), 201–216. <https://doi.org/10.1080/15391523.2007.10782505>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.