CS-411 Project
Muscular Reinforcement : Learning Burpees

Group H : Tobia Albergoni, Matteo Yann Feo, Diana Petrescu

January 15, 2019

1 Introduction

The purpose of this document is to describe the design process, execution and analysis of a small learning activity prepared in the context of the EPFL course *CS-411 Digital Education and Learning Analytics*. This learning activity takes the form of a 25-30 minutes lecture that aims at conveying to the students information about the *Burpee*.
A *Burpee*, sometimes also called *Squat thrust*, is an aerobic exercise that involves both upper and lower body muscles and that can be used for a fairly complete strength training workout. The lesson teaches to the participants the basics of proper and correct muscular training with an engaging and hands-on approach, aiming to provide the necessary knowledge to allow them to perform the *Burpee* in a safe and efficient manner during their workout sessions.
The project consists of a design process that tries to adopt some good practices of learning sciences theory and conform to established pedagogical models. The resulting lecture is then carried out during experiment sessions with groups of participants, where data is collected and subsequently analyzed. The analysis aims at assessing the learning outcomes of the activity and answer some research questions that will be presented in later stages of this document.

1.1 Motivations

Workouts and muscular reinforcement exercises are performed everyday by a large amount of people, with the intent of getting or staying in shape. A correct and repeated execution of this kind of exercises leads to a healthy and performing body. However, doing exercises incorrectly could do more harm than good and cause both short and long term injuries, and it's therefore important to know the theoretical and practical background for each training exercise. Spreading knowledge about correct exercising practices is beneficial for the overall well-being of people, which is one of the main reasons for choosing this topic as the scope for the project.
More specifically, the *Burpee* has been identified as an appropriate topic for this project because it can easily be decomposed in multiple parts and taught incrementally, and combines both practical and theoretical knowledge. This intrinsic modularity and multi-dimensionality provides an ideal context to design a sufficiently complex and pedagogically sound learning activity.

2 Objectives

As a good instructional design practice, the objectives and expected outcomes of the activity are first briefly presented and discussed before touching on the actual lecture design. The following subsections clarify what the students should be able to do at the end of the lesson.

2.1 Learning goals

The main learning goals of the lecture can be divided into a couple of categories.
First and foremost, by the end of the lecture participants should possess the necessary knowledge to correctly describe a *Burpee* in its entirety, including the details of each sub-exercise. This includes knowledge about positioning of body parts, angles, direction/speed/range of movements and sequencing of all steps. This set of information should leave the student with the necessary tools to perform and explain a correct *Burpee*, improving quality of training and reducing the risk factors derived from incorrect executions.
Secondly, the participants should acquire basic knowledge about the body parts and muscles involved in each exercise step. They should be able to briefly motivate the reasons for doing/avoiding certain movements or postures. This includes understanding incorrect actions and the relative risk factors.
Third, all of the above knowledge should be at least partially transferable to muscular reinforcement exercises in general. Students should hence be able to take some of the information learned in the context of the *Burpee* and apply it to similar workout exercises.

Lastly, the acquisition of practical experience in doing a correct *Burpee* is a complementary goal of the lecture. Given the short nature of the course and the heterogeneity of participants' physical shape and prowess, it is not expected that a student is able to master the exercise in just a handful of repetitions. This component of the learning activity can nevertheless be formulated as having the goal of giving an intuition about practical execution of the *Burpee*.

2.2 Pedagogical objectives

After having clarified the general context-specific goals of the lecture, the educational learning objectives are framed more formally by referring to Bloom's taxonomy. The purpose of the activity is to reach mainly the following cognitive levels:

Remember The different steps of the *Burpee*, the correct postures, angles and movements, the muscles involved and the associated risk factors.

Understand The motivation behind postures/movements and the effect of the different body parts on the exercise.

Apply The knowledge relative to the *Burpee* and its sub-steps to other muscular reinforcement exercises by arguing about similarities and dissimilarities.

This is the set of cognitive sub-domains commonly reached and tested (and arguably over-represented) by classical education. A participant to the lecture is expected to need the information provided during the activity primarily in situations where he has to either replicate the exercise, explain it to someone, or perform another workout exercise that involves similar body parts. Therefore, it is not deemed beneficial to try to reach more deep cognitive levels in this project, since they do not fit the context particularly well and would require more complex and gimmicky activities that are not easily included in such a short course.

2.3 Lecture outcomes

The objectives section is concluded by list of more concrete expected outcomes. Listing them explicitly goes in the direction of *writing the exam before the course* and pinpoints how students should be evaluated at the end of the activity. The list anticipates some of the content sub-divisions that will be clarified during content analysis.

- Students should know that a *Burpee* is composed by a *Squat*, a *Pushup* and other chaining movements.
- Students should be able to explain the execution step by step for each sub-exercise.
- Students should be able to explain the correct sequencing of steps across the whole *Burpee*.
- Students should be able to mention important postures and movements for each sub-exercise.
- Students should be able to identify wrong executions of the *Burpee* or its sub-exercises by looking at a video, picture or text description.
- Students should know which body parts are at risk as a consequence of wrong executions.
- Students should be able to name the muscles involved in each sub-exercise and their relevance/role.
- Students should know some variants of the sub-exercises and explain the difference with the original version.
- Students should be able to get in the correct starting position for the *Burpee* and each sub-exercise and begin the correct execution.
- Students should be able to evaluate a live performance of a *Burpee*, identifying mistakes and giving advice.
- Students should be able to analyze other exercises and identify and argue about similarities and dissimilarities with the *Burpee* and its sub-exercises.

3 Prerequisites and target audience

3.1 Prerequisites

Considering the objectives set by the previous section, it is straightforward to come up with a list of prerequisites that participants should possess before starting the course.

Health Above all, some preconditions on the physical state of the students should be set in order to guarantee that the practical parts of the lecture can be executed with no health-related risks for anyone. As the activity includes trying both the individual sub-exercises and a full *Burpee*, injured or physically debilitated individuals should avoid doing it. Alternatively, such people could follow the theoretical explanations and restrain from doing the practical parts, but this would cause a non-uniform experience and data collection across participants. Therefore, being healthy is a strictly enforced prerequisite in the context of this project.

Age Furthermore, targeted muscular reinforcement is advised starting from the age of 12/13 onward, when the body starts developing towards that of an adult. On the other end of the spectrum, the *Burpee* is a dynamic and explosive exercise, so old people should be careful when attempting it. This means that the course should be taken only by teenagers and adults whose body is suitable for muscular workout.

Prior Knowledge The content of the lecture is quite straightforward. It does not require a given level of education or particular context-specific knowledge, since everything is explained from the basics. Nevertheless, some muscles and body parts are mentioned during the course, hence some rudimentary anatomical knowledge can be useful to avoid relying solely on the graphical indications shown during the lesson. The learning activity requires fluency in the English language.

3.2 Target audience

Given that the lecture does not have demanding requisites in terms of theoretical background or previous experience, it should appeal to a large audience of young adults and adults looking to improve the efficacy and correctness of their training. Given that the experiment participants are likely to be fellow university students, they belong to the intended target audience.

4 Content analysis

In this section, the *Burpee* is decomposed and disentangled into its core components and laid out in a concept map, which will serve as a guideline for the actual lecture design. Given the intrinsic importance that visuals, videos and images hold during the explanation of a physical exercise, only a brief recapitulation of the most important points is presented here, while the lesson will expand on those points with images, schemes and videos.

4.1 Map of a generic muscular reinforcement exercise

First, since the content analysis is about a training exercise broken down into sub-exercises, we present a content template for a general muscular reinforcement exercise, that abides to the aforementioned learning goals. Each *Burpee* step will be analyzed according to this template.

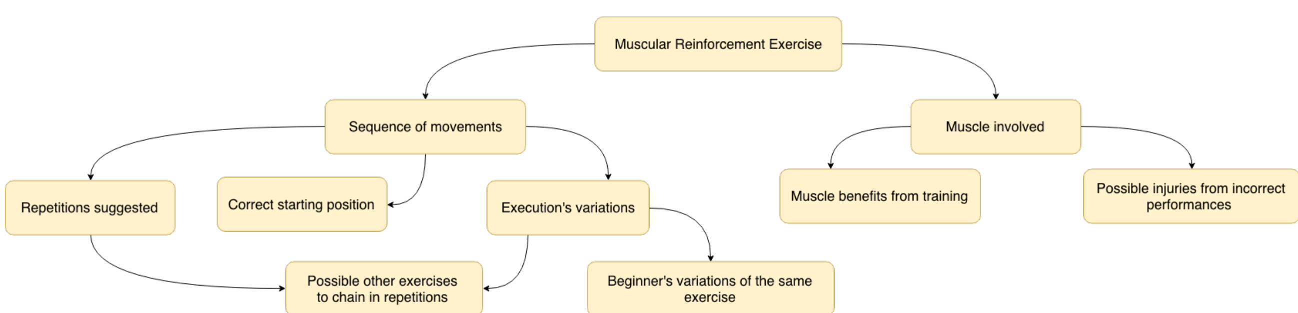

Figure 1: Content analysis of a generic muscular reinforcement exercise

4.2 Components of a Burpee

It was already mentioned that the *Burpee* presents the advantage of being easily decomposed into multiple easier steps and sub-exercises. This proves to be useful for analyzing the content and structuring the lecture. The following sequence of steps is the variant of the *Burpee* considered in this project:

1. Begin in standing position
2. Move into *Squat* position
3. Hands on the ground and kick feet backwards
4. Perform a *Pushup*
5. Return feet into *Squat* position
6. Stand up from *Squat* position
7. Jump and resume standing position

This sequence makes it clear that the *Burpee* is built upon two simpler exercises, the *Squat* and the *Pushup*, chained together with some additional movements. Those three components are now briefly presented.

4.2.1 Squat [4] [6]

The *Squat* is a widely known exercise for upper leg muscles training.
To perform a squat, the person should start in standing position, with the feet a bit wider than shoulder width and pointing slightly outward to the sides. Intermediate students could keep them pointing towards the front, as long as they can keep the rest of the exercise under complete control. Arms can be kept extended straight forward or crossed forward in order to help keep balance.
The movement consists in lowering the whole upper body by bending the knees and descending, then ascending again until reaching starting position. During the whole movement, back should remain straight and abs should be contracted. Body weight stays on the heels for the entire exercise. The descent should go as low as possible while still keeping shins vertical and heels on the ground. Pay attention not to pull knees while squatting and keep them pointing aligned with feet. Knee tip should never extend beyond the toe.
The primary muscles used during a *Squat* are two large muscle groups in the thighs: *Quadriceps* and *Hamstrings*. Multiple other leg, back and gluteus muscles support the movement and help keep balance.
Going too low during descent can cause a retro-version of the pelvis, which could be very damaging for the back. Additionally, feeling pain in the knees at any moment during the movement is and indicator that the exercise is not being performed correctly and that articulations are over-stressed.

Figure 2: Correct starting position to perform a squat (left)
Descending phase of the squat (middle)
Muscles involved during s squat (right)

4.2.2 Pushup [8] [2]

The *Pushup* is another popular workout exercises that focuses instead on upper body strength.
A proper *Pushup* is executed by starting in prone position, but with the body kept afar from the floor by the extended arms and by standing on the toe tips. The wrists should be placed approximately in a vertical line under the shoulders. The wideness of the arms influences which parts of the muscles are trained, but both wide and narrow stances are correct. The whole body should be kept in a straight line, from feet to the head, with spine and legs perfectly aligned. Neither the butt nor the back should form inward or outward arches. In order to help keeping the body straight, the pelvis should be rotated backwards and the buttocks squeezed while keeping the abs contracted.
One repetition involves lowering the whole body towards the floor at once by contracting the arms, and returning back to the starting position by extending them. The full range of motion should be performed in order to maximize efficacy. While contracting and bending the arms, pay particular attention to the direction of the elbows: they should point backwards and absolutely not sideways. Additionally, shoulder-blade movement has to be enforced, pulling them inwards when descending and then pushing them apart when returning back up.
The *Pushup* involves primarily the *Pectoralis Major*, the *Triceps Brachii* and the *Anterior Deltoids*, on top of other stabilizer muscles such as abs or lower back muscles.
The exercise should be performed slowly and controlled, since quality of repetitions is better than quantity. Having the elbows point sideways during descent is a major risk factor and could lead to severe shoulder injuries.

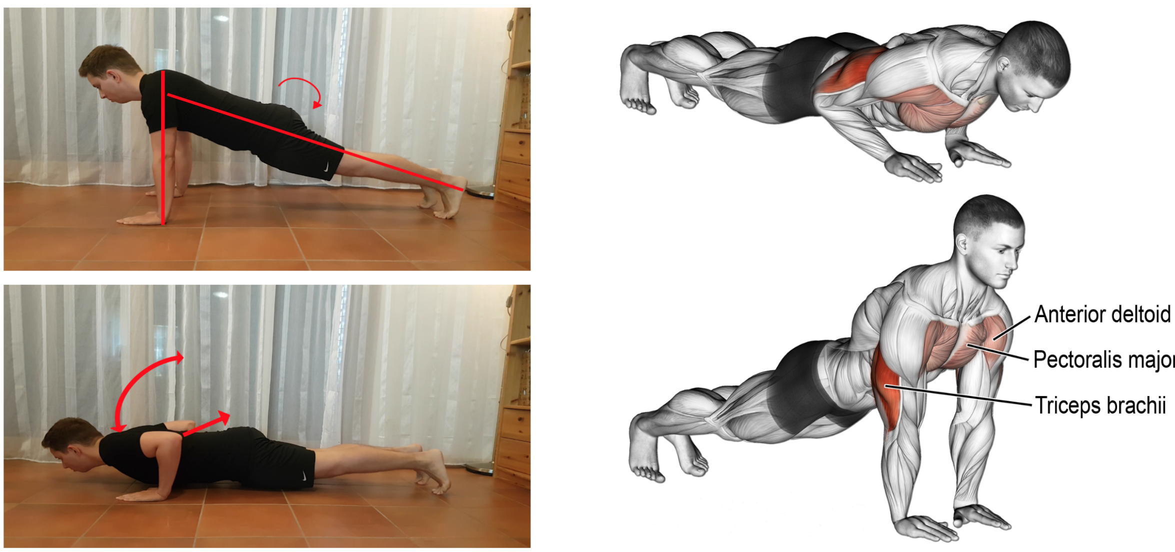

Figure 3: Correct execution of a pushup (left)
Muscles involved during a pushup (right)

4.2.3 Chaining movements [1] [7]

In order to produce a *Burpee* from the two above exercises, one has to perform a series of chaining movements that allow to transition between them.
The exercise is started in the starting *Squat* position. A full *Squat* descent is performed, but instead of returning back up, hands are put on the floor in front of you at *Pushup* distance. At this moment, feet are jumped backwards and the body should land in the starting *Pushup* position. A full *Pushup* is then performed until the body is back in planck position, at which moment an opposite jump with the feet is done and hands are raised from the floor. The body should find itself in the lowered squatting position, from which the ascending part of a *Squat* is performed. The ascension is done faster and with more energy, such that when the legs are completely extended a small jump is done with the arms raised over the head. The landing should bring you back into the starting position, marking the end of one *Burpee* repetition.
It's important to keep balance during each chaining movement and make sure to land in a proper exercise starting position. Correct the position if necessary, but don't execute a *Squat* or *Pushup* incorrectly as a consequence of bad chaining. Additionally, hard landings should be avoided when jumping or moving feet, and the head should always follow the movements naturally facing forward to avoid neck pains.

4.2.4 Other

To conclude the actual lecture content, a few indications to perform muscular reinforcement in the correct way. First, some warm-up should be done before starting to exercise, for example with a short run. This can prevent joints and muscles injuries. Second, every movement should first be done slowly and in a controlled way. Speed can be added when the exercise is mastered. Quality is much more important than quality and is a safer way to exercise.

4.3 Lecture content map

Finally, all of the above information can be inserted into the exercise schema template presented above and tied together following the correct sequence of sub-exercises that form the *Burpee*. The resulting complete content map of the lecture can be inspected in the following figure.

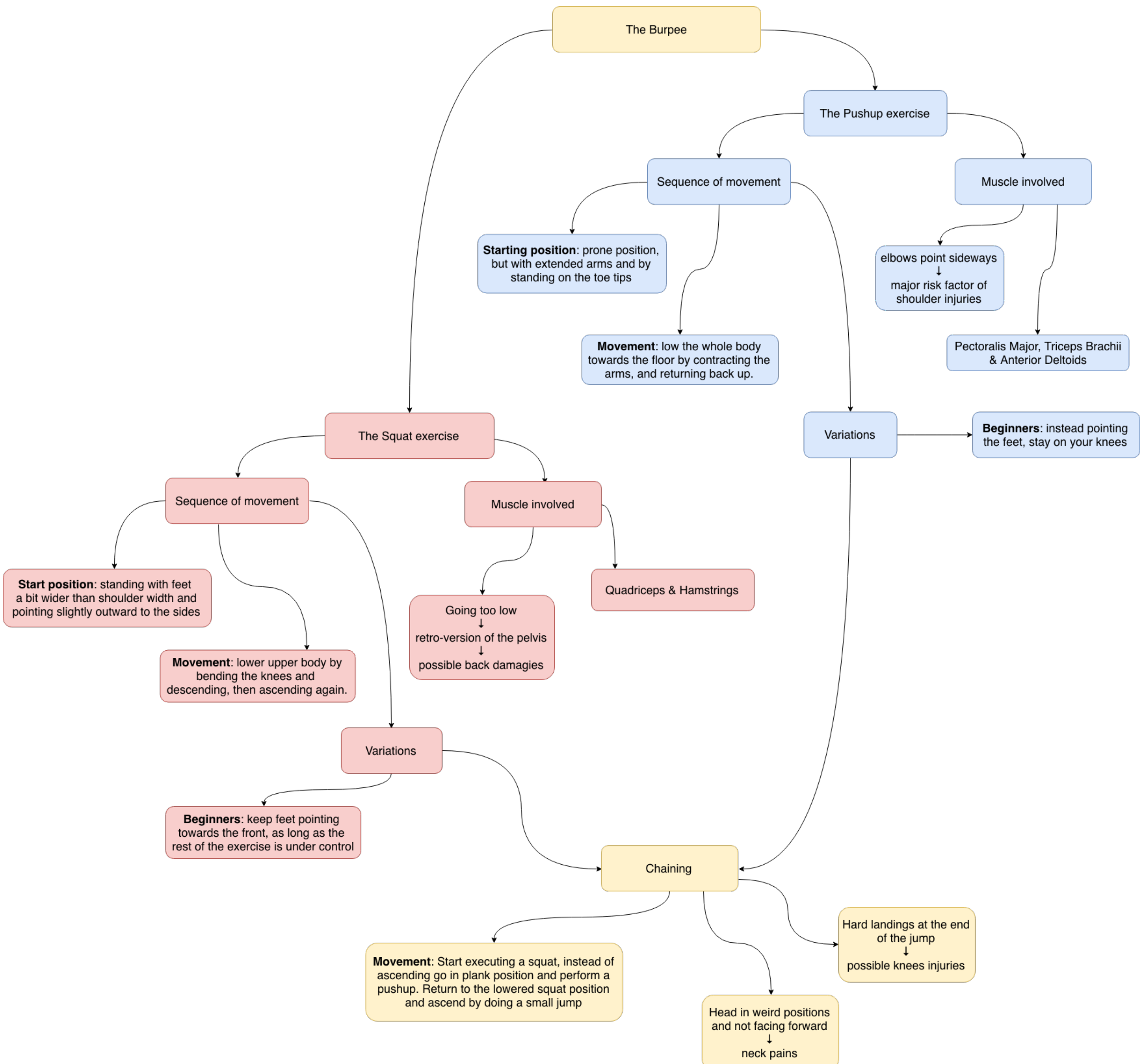

Figure 4: Lecture content map of the Burpee exercise

5 Lesson design

In this section, the motivations and choices regarding actual lecture design are presented. First, the high-level design choices are clarified and motivated by course content. Then, each part of the Orchestration Graph is briefly presented. This section builds upon what was introduced in the objectives and content analysis parts of this document in a straightforward way.

5.1 Pedagogical models

5.1.1 Frame based model

As explained during content analysis, the *Burpee* is easily decomposed into sub-exercises that can then be chained together. This makes it suitable to be taught incrementally, with each sub-exercise being a building block that can be learned, tried and evaluated in isolation. This is the reason why a frame-based model is chosen for this topic: the course is split into *Squat*, *Pushup* and *Chaining* modules. Each module contains a theory part, where the information regarding the sub-exercise is given by the instructor, a practical part, where the student can try the exercise individually, and an intermediate test about the contents of the module. This intermediate test provides immediate feedback about the answers. Such a module abides to the characteristics of a frame-based learning activity as there is obviously decomposition of content, the student is kept active and involved at all stages through quizzes and practical parts and the course provides immediate feedback to students through quiz answers. There is also a degree of individualization, since the theory parts are implemented through videos, which will be presented just below.

5.1.2 Socio-cultural elements

Taking inspiration from the MOOC approach, the project pairs the above frame-based model with some activities and components that introduce collaborative and social elements.
First, for all experiment sessions, the practical part of the last module (*Chaining*) is done together with the whole class. Each participant attempts some *Burpee* repetitions while the other students watch. After the execution, each student has to grade the performance of the others and give some personal feedback. This peer evaluation has the advantage of providing a feedback which should be more likely to be internalized, as it comes from fellow students and not from the instructors. Hence, the verbal interactions should remain in the *Zone of Proximal Developement*, as the guidance on exercise execution comes from peers that are, at best, just slightly more knowledgeable on the topic.
The second and arguably much more impactful collaborative element is introduced in the course as the controlled independent variable of the experiment. Half of the participants will make up the experimental group and follow a modified version of the lecture. In this more collaborative version, participants are paired by a grouping operator during the practical parts and the intermediate test of each module. The students have the opportunity to try the exercises with another person instead of individually and exchange feedback. Also when returning back to the computer to answer the intermediate quizzes, they will have the opportunity to chat with the student they were paired with and exchange opinions on the questions and answers. The answers to questions are still individual, the students can choose to disagree after discussion. This pairing component gives the opportunity to verbalize with a peer all the course content, something that the control group cannot do until the final practical part.

5.2 Technologies and learning mediums

The lecture is obviously implemented using the FROG framework, which is not explained in this document, but introduces in the Orchestration Graph some components that need to be presented, as they lie outside the common FROG context.

5.2.1 MOOC studio videos

All the information provided by the instructors is presented to the students through videos. The videos were registered at the EPFL MOOC studio with professional equipment. They consist mostly in the instructors talking about course content, supported by images and short clips. The videos, given the choice of a frame-based model for the course, provide a much needed degree of individualization, since the students can process new information at their own speed and can re-watch unclear parts before moving on with the rest of the course. This possibility is somewhat limited during experiment sessions, since there are time constraints to be respected. But if a student were to take the course outside the scope of this project, he would have the freedom to go at his own pace.

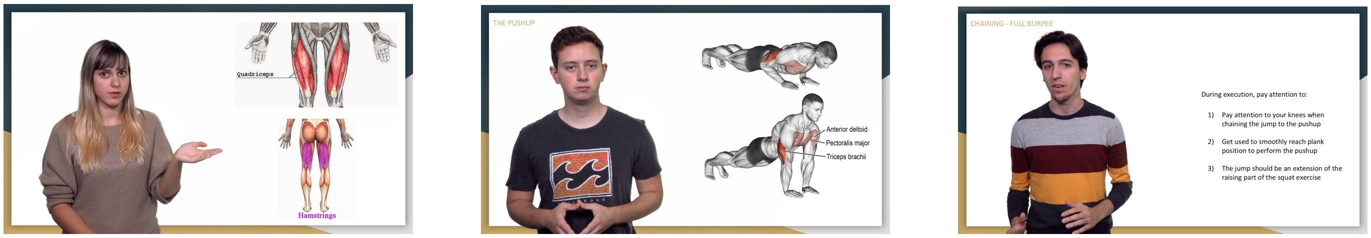

Figure 5: Frames from the different videos recorded

5.2.2 Google Forms

Except the pre-test, which is implemented directly with the appropriate FROG activity, all tests of the lecture are administered through Google Forms, embedded in FROG using the right activity. This choice is made primarily because, in order to comply to the aforementioned model, quizzes need to provide immediate feedback after each answer, something that Google Forms allow to do without the need to modify the FROG quiz activity at the code level.

5.3 Structure and sequencing

The concrete activities that form the OG are now presented in detail. The reasons for their presence in the course is clarified if they are not already motivated by the high-level pedagogical decisions mentioned above.

5.3.1 Introduction and Pre-test

During the introduction, a short video is shown to the participant where the lecturers explain the goals of the lesson and present a complete *Burpee*, while highlighting the benefits of such exercise. This is done in order to hopefully provide an intrinsic motivator for the class. This is done right at the beginning to captivate the students for the rest of the lecture.
This video is immediately followed by a quiz that serves as a pre-test and allows to inspect the familiarity of participants with the *Burpee* and with muscular reinforcement in general. Pre-representations and general knowledge about workouts is tested during this phase. The collected data is not used during the rest of the lecture but is important for the subsequent data analysis in order to assess biases among experiment groups.

5.3.2 Modules: The Squat and the Pushup

The first two modules of the course are identical in terms of structure. They are naturally pre-requisites for subsequent activities, as they are the building blocks of the *Burpee*.
First, a short video explaining the exercise is presented to the participants. The video contains all the theoretical information presented in the content analysis section, supported by images, videos and diagrams. Second, a blank FROG activity leaves the participants some time to choose a corner of the room and try to perform the presented exercise. This activity should force the students to proceduralize the information given in the theory videos.
Then, an intermediate quiz about the notions is proposed, where the answers to questions appear right after each response with a complete explanation, in order to provide precious immediate feedback. These quizzes test for retention and understanding of the information, while analysis questions that require a higher cognitive level are left for the final test of the lecture.
Recall that it is in this part of the course that the experimental group is paired and has the opportunity to perform collaborative learning. The grouping operator pairs participants at random.

5.3.3 Last module: Chaining movements

The last module of the lecture consists only in a video clarifying how the two previous modules should be chained together to produce a *Burpee*. Since the chaining movements cannot be properly tested in isolation, this module does not have a practical part nor a quiz, and its contents are tested in the last practical activity and in the final test.

5.3.4 Practical part and Peer evaluation

After the *Chaining* theory module, the class is gathered in the open space and each participant is asked to perform some *Burpee* repetitions. The rest of the class watches the performance and has to grade it. Note that the class is notified at the beginning of the lecture of this activity, in order to allow even shy people to mentally prepare for a "public performance". This practical part puts some pressure on the participants since the beginning of the course, which has to be followed with commitment in order to be able to show a proper *Burpee* at the end. Apart from providing social interaction, already discussed above, this activity produces interesting data for the subsequent analysis in the form of peer grades.

5.3.5 Final test

The lecture is concluded by a last test, that covers the material of all three modules. This quiz does not provide feedback after each question and, most importantly, does not limit itself to the cognitive levels of remembering and understanding. Some questions ask the students to analyze and reason about a different muscular reinforcement exercise and identify similarities or differences with the *Burpee* and its sub-exercises. This activity introduces a knowledge transfer component in the lecture and is the principal indicator of learning outcomes.

5.4 Orchestration Graph

All the aforementioned activities can be found in the right order in the following Orchestration Graph [3]. Note that this figure represents the lecture version where participants are paired to discuss answers to quizzes. The individual version is identical, but without chat activities and the grouping operator.

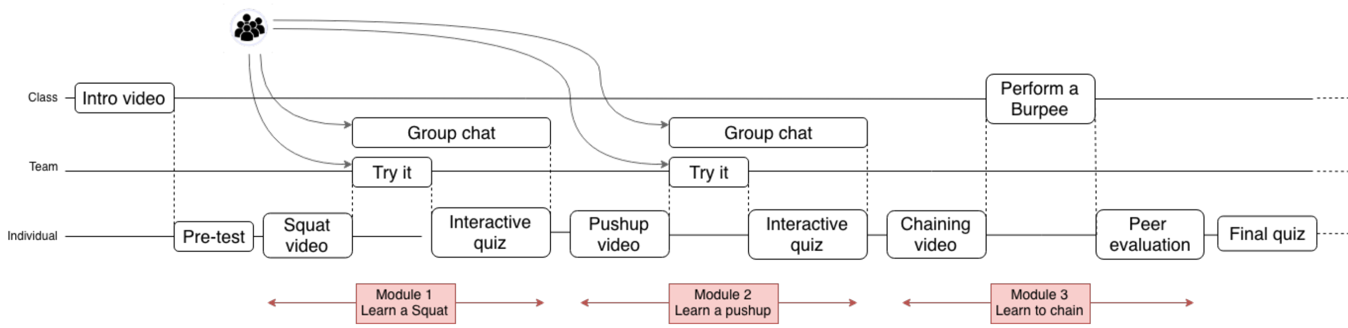

Figure 6: Orchestration graph of the group lecture

6 Experiment design

This section clarifies how the lecture was presented to students in the context of this project and how the experimental setting was devised and put in place.

6.1 Class composition and logistics

Due to the space requirements needed in order to allow the participants to execute the different exercises without risks, the decision is to limit the class size to only 4 people and to conduct multiple experiment sessions. This allows the instructors to closely monitor students behavior and actions and make sure everyone follows the course in the correct way and that all intended data is collected. This approach additionally presents the advantage that the data collection process is incremental, and if results are not particularly strong or clear with a certain data-set size, it's easy to plan new sessions and obtain more data points in reasonable chunks. There is no fixed setting for the experiment sessions in terms of location. It suffices that the place has a table where 5 computer can run the Orchestration Graph (1 instructor + 4 students) and enough space to perform *Squats*, *Pushups* and *Burpees*.

6.2 Independent variable

As explained and motivated in the lecture design section, the independent variable of choice for this project consists in having half of the participants do some parts of the course in pairs instead of alone. Concretely, in each module they will practice the exercise together and have the possibility to chat during the intermediate tests, even though the answers are still individual. The choice is to have two separate Orchestration Graphs and perform entire experiment sessions with either the paired or individual variant. Consequently the number of sessions must be in equilibrium between the two variants in order to have experimental and control groups of the same size.

6.3 Data collection

We now clarify which data is collected during an experiment. First, recall that all the quizzes except the pre-test are administered to the participants through Google Forms embedded in FROG. Hence, at the end of a session, the forms answers are exported in CSV format. This covers the answers to the two intermediate module tests (*Squat* and *Pushup* tests), the peer grades and feedbacks and the answers to the final test. The rest of the data (pre-quiz answers and chat data in the case of paired version) is collected by exporting the FROG session at the end of the experiment, which provides the data in JSON format.

7 Analysis

This section describes the analysis conducted on the data that was collected during the experiments, from questions to data processing, concluding with answers and interesting results.

7.1 Research questions

The analysis should start from the interrogatives that the project is trying to answer. The research questions are hence first presented and briefly motivated.

Q1 Are the desired learning goals attained by the participants? Does the frame-based approach experimented in this project allow to teach muscular reinforcement exercises?

Q2 How do the learning gains evolve across the different tests of the lecture?

Q3 Is the collaborative learning approach effective in this context? Did the introduction of pairing as an independent variable have an effect on the learning outcomes?

Q4 After the lecture, are the participants able to transfer the knowledge about *Burpees* to other muscular reinforcement exercises? To which extent?

For what concerns the first research question, the intention is to measure the scores of participants in the final test and evaluate them based on the expected learning outcomes set at the beginning of this document. This is in general the primary result that the analysis should pursue: have the students reached the goals set during lecture design? This should allow to argue whether or not the learning activity design is suitable for the chosen topic.
Given that the frame-based approach with intermediate module tests provides feedback about student retention across the duration of the small course, it might be interesting to look at the evolution of all test scores, from pre-test to the final quiz. This should be done while keeping in mind that the different quizzes target different cognitive levels.
Third, the effects of the independent variable introduced in the experiment have to be assessed. In this case, the independent variable introduces pairing and collaboration during the lecture. Observing and comparing test scored between control and experimental groups should allow to assess whether the collaboration between participants was beneficial or not.
Finally, as teaching one single exercise is a very specific lecture goal, it is important to examine whether or not the knowledge provided in such a restricted context can be transferred to and generalized to muscular reinforcement. In order to do this, the scores obtained by participants in questions that evaluate transfer learning should be analyzed.

7.2 Data processing

7.2.1 Data

Over the course of the project, 6 experiments of 4 participants each were conducted, for a total of 24 participants. Three sessions were done with the *individual version* of the Orchestration Graph and make up the control group of 12 people. Similarly, the remaining three sessions were conducted with the *grouped version* of the OG.
For each participant, the following data is available. First, the answers to the 3 pre-test questions. Second, 4 answers for each of the two intermediate module tests (*Squat* and *Pushup*). Third, 12 answers to the final test. Additionally, more data is available in the form of 3 peer grades/evaluations per participant, from the rest of his group.

7.2.2 Tools and frameworks

The data was processed and analyzed with the *Python 3* programming language. More specifically, the following packages proved to be useful to conduct the analysis: *pandas* for loading and manipulating the data, *scipy* for statistical analysis and *matplotlib* and *seaborn* to plot results.
The analysis is conducted with the *Jupyter Notebook* framework.

7.2.3 Processing

As quiz answers are stored differently (JSON for pre-test, CSV for the rest) and consist in various answer formats (multiple choice answers, single choice answers, textual answers), data needs to be cleaned up before analysis. Quiz answers were approached as follows: each question was designed to have the same weight in scoring, and a value between 0.0 and 1.0 was thus assigned depending on the correctness of the answer. For binary questions, score is either 0.0 or 1.0, for multiple choice the proportion of correct checks is the score, and for textual answers a value is assigned by the instructors depending on the quality and argumentation. This allows to have a uniform scoring mechanism across the entire data-set and to compute the percentage of obtained points for each quiz for all participants. This processed data is the basis for most of the analysis.

7.3 Analysis and results

7.3.1 Achievement of expected learning outcomes

The final quiz is designed to be challenging and touch on three different cognitive levels. Reaching the learning goals doesn't therefore mean obtaining a perfect score. The following "grading scheme" is instead used as a reference in order to evaluate the attainment of the objectives:

Excellent	**Very Good**	**Good**	**Satisfactory**	**Sufficient**	**Insufficient**
100%	90%	80%	70%	60%	<60%

Table 1: "Grading scheme"

The following figure and table summarize the final scores of the participants.

Students	**Mean**	**Std**	**Median**	**Min**	**Max**
24	85.32%	7.64%	86.95%	62.60%	94.86%

Table 2: Final test scores descriptive statistics

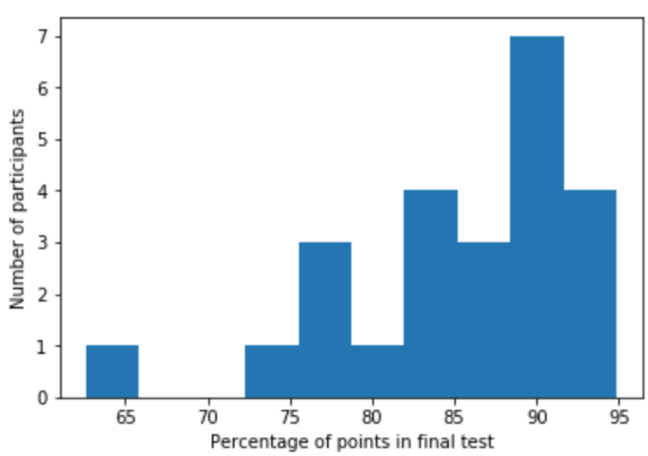

Figure 7: Distribution of final test scores

It can be noted that both the average and the median lie in the *good/very good* range. Additionally, no participant has an *insufficient* score, with just one outlier scoring a minimum of 62% and the rest being above 70%. According to the quiz design and the "grading scheme" these results allow to conclude that the students end the lecture with a knowledge level that satisfies the expected learning outcomes.

7.3.2 Evolution of quiz scores

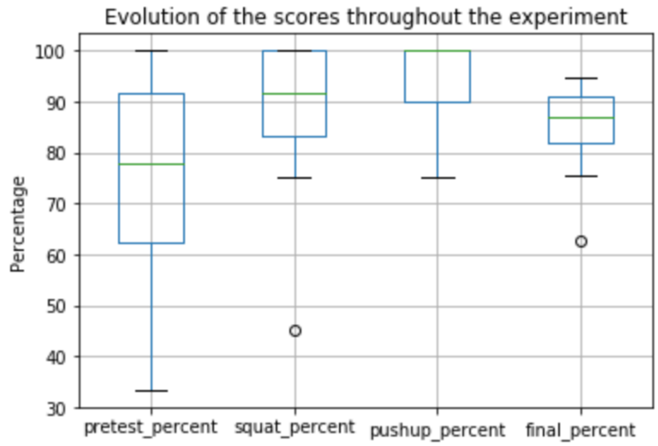

Figure 8: Evolution of test scores

Pre-test	Squat test	Pushup test	Final test
74.77%	88.54%	94.79%	85.32%

Table 3: Means of test scores

In the above figure the evolution of participants average score across the different lecture tests is displayed thanks to a box-plot for each test, and the mean percentage for each test is clarified in the subsequent table.
At the end of this section, the inconclusiveness of the pre-test data will be showcased and discussed, hence it's wiser not to include its score in this analysis. Focusing on the intermediate tests mean scores, it can be seen that both lie in the *Very good/Excellent* range, with participants slightly improving in the *Pushup* test, albeit not in a particularly meaningful way. This could be motivated by the fact that students are a bit more accustomed to FROG and the quiz structure when they complete the second intermediate test, hence they perform a bit better, but we cannot prove this conclusively. Nevertheless, the generally high scores in both tests demonstrate that students manage to retain most of the theory video information immediately after watching them and are able to remember and understand the two sub-exercises pretty well.
The final test mean score is slightly lower that intermediate tests, but that should not be a surprise for different reasons. First, the final test is designed to be longer and harder than intermediate tests. Second, it also includes transfer learning evaluation, which touches a cognitive level (*Application*) that is not reached with the other quizzes. Third, it evaluates the students 10-20 minutes after they received the information and after practical activities that are likely to shift their attention to more procedural knowledge. Therefore, this lower score in the final test should not be considered a particularly surprising or meaningful result.
This overview on test scores over the whole experiment session highlights that students results are also strong in terms of immediate retention and understanding, on top of the final learning outcomes discussed in the previous sub-section.

7.3.3 Impact of independent variable

The analysis of the impact of collaboration on learning is conducted in two ways: first, the mean final scores for control and experimental groups are compared and discussed. Second, it's deemed interesting to inspect the *agreement rate* of experimental group pairs on intermediate quiz answers compared to the agreement rate of random pairs in the control group.

Individual (Control)	Paired (Experimental)
81.24%	89.39%

Table 4: Mean final test scores for control and experimental groups

In order to compare the final score distributions for the two groups, a one-tailed t-test for two independent samples with non-equal variances is used. This test has a normality assumption on the two samples, which has to be formally confirmed. In order to do so, a *Shapiro-Wilk [5]* test is performed on both samples, confirming the normality assumption for both the control ($p = 0.23$) and experimental ($p = 0.09$) groups, since the null hypothesis cannot be rejected. This result is confirmed by QQ-plots, not reported here for the sake of conciseness. The t-test can hence be applied to the samples and results in a p-value of 0.007, which allows to conclude that the final scores of the experimental group are meaningfully higher than those in the control group. This is an interesting result, as it seems to show that adding a collaborative component to the lecture leads to much better final results. This is probably motivated by the fact that having to verbalize all the theory in a chat with a partner improves understanding and retention.

The *agreement rate* is instead computed as follows. For the experimental group, the percentage of matching answers to intermediate test questions (the two test that the pairs did collaboratively) of the two members of the couple is computed. Recall that participants are not required to give the same answer even if they are paired, hence it can be interesting to see the degree of actual agreement. In order to compare the resulting percentages to the control group, the latter is randomly grouped into pairs, and the *agreement rate* of those random pairs is computed in the same way. This procedure is done repeatedly for the control group, to ensure that the randomness of pairing gives stable results. The resulting *average agreement rates* for the two groups can be inspected in the below table.

Individual (Control)	**Paired (Experimental)**
75.00%	83.33%

Table 5: Mean agreement rates on intermediate quiz answers for the two groups

The hypothesis is that the degree of agreement is higher for the experimental group as a consequence of the collaboration in the chat activity. This is tested with the same procedure used for mean final scores, but since the normality assumption on the experimental sample does not hold, the Kolmogorov-Smirnov statistic is computed instead of using a t-test, and should allow to determine the likelihood of the two samples originating from the same distribution. Sadly, a p-value of 0.43 doesn't permit to reject the hypothesis that the underlying distribution is the same. Therefore, no meaningful conclusion on the *agreement rate* can be drawn from collected data, and further experiments are needed to gain insights on this area.

7.3.4 Transfer learning

In order to analyze the degree of knowledge transfer that the participants are able to reach by the end of the lecture, the final test questions that measure transfer learning are isolated and analyzed. The participants overall achieved a *Good/Very good* average result of 84.4% of the points in the transfer learning questions. This alone confirms that participants are able to apply the knowledge structures learned in the *Burpee* context and apply them to other exercises. It has to be said that the final quiz challenged students with near-transfer problems where the answers were not particularly difficult to guess. Nevertheless, since the participants were asked to motivate their answers in textual form, the good quality of answers can be confirmed and allows to conclude that students are indeed capable of using the new knowledge outside the *Burpee* context.

It's interesting to inspect what types of students managed to perform well in the transfer learning questions. In order to do this, the Pearson correlation coefficient is computed between transfer learning scores and final test overall scores, which yields a positive correlation value of 0.84, which is quite high. This seems to hint at the fact that students that managed to master the information provided by the lecture were also those that were able to transfer this knowledge to other exercises, while those who did less good in knowledge transfer also had troubles understanding the rest of the material. This might seem a trivial result, but it may confirm that the students that really understood the course content were the ones able to use it at deeper cognitive levels. The scatter-plot in figure 9 shows the existence of this linear correlation graphically.

7.3.5 Inconclusive data

The last part of the analysis and results section is reserved for the presentation of data and results that were designed to provide some insights, but finally proved to be rather inconclusive. One aspect that produced one such situation is the pre-test.

The pre-test was designed to provide insight about the level of prior knowledge of students about the *Burpee* and muscular reinforcement in general. Therefore, some correlation between pre-test scores and overall performance during the lecture was expected. The scatter-plot of figure 10 shows the pre-test scores against the final test scores. It's rather evident that the two are not particularly correlated, a guess confirmed by low coefficient values both for Pearson's r (0.19) and for Spearman's *rho* (0.066).

The fact that there is no correlation whatsoever between final scores and a measure that should indicate the degree of prior knowledge raises interrogatives on the accuracy of the latter. After reviewing the pre-test questions with this

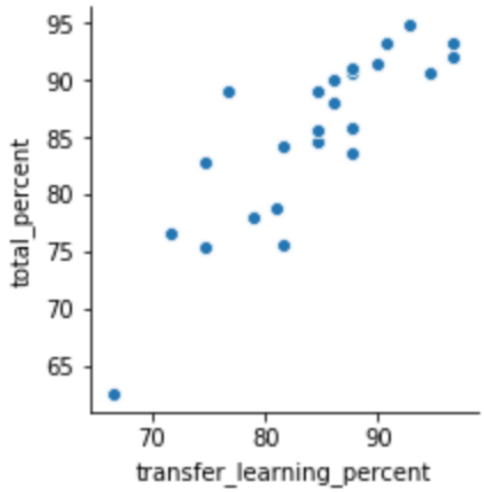

Figure 9: Scatter plot of transfer learning scores and final scores

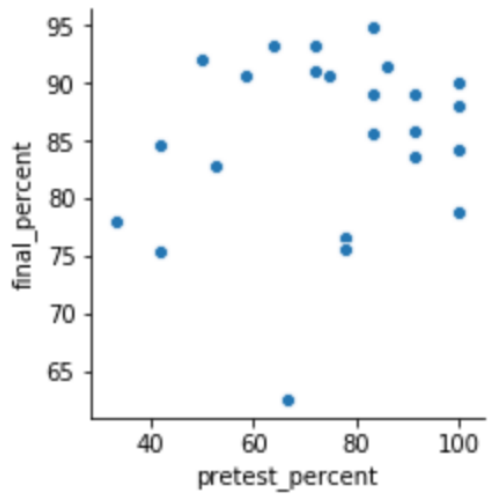

Figure 10: Scatter plot of pre-test scores and final scores

result in mind, it was realized that the questions were too general and not tied enough to the learning goals to provide a value usable in comparisons. Hence the pre-test results were not used in order to draw meaningful results, as the confidence in their reliability was greatly reduced.
Another aspect of the data that was completely disregarded during the analysis were the peer evaluations that students gave to other participants. Although in principle that would have been an additional measure of the learning outcomes, in reality the project failed to set proper guidelines for the numerical peer evaluation, which resulted in highly inconsistent grading criteria across the different experiment sessions. Overly kind or harsh students caused the data to be a really unreliable performance score, and the decision was hence to avoid using it to draw results, similarly to pre-test scores.

8 Conclusions and evaluation

8.1 Project evaluation

Overall, the project was able to produce a cohesive and well-structured learning activity that the participants seemed to enjoy. Based on the experimental design presented in this document, the learning goals set at the beginning of the project were attained with seemingly high success, which is a huge positive in the evaluation of this work. The OG was designed in order to collect a lot of different data, which proved to be a considerate choice in order to obtain meaningful results even with some activity design errors. Such errors consisted mainly in the design of the pre-test and in the introduction of peer evaluation in this format. The ideas behind those two concepts were pedagogically sound, but their execution was lackluster and they therefore did not contribute to project results. On the positive side we can instead find the choice of a collaborative component as the independent variable, as it yielded clear results. On top of this, the recording of MOOC style videos was a fun and interesting experience, which produced high quality educative multimedia content. In general, learning how to think, analyze, decompose, sequence, design and execute a learning scenario was an enriching experience.

8.2 Conclusions

To conclude this report, the main results of this project are summarized. First, an inherently practical and physical activity like muscular reinforcement seems to be easily taught with digital supports such as the FROG framework and video lectures. Additionally, a frame-based pedagogical model appears to work well in order to present complex exercises with multiple sub-steps, to the extent that students are able to transfer exercise-specific knowledge to other contexts. All these claims are supported by the high degree of satisfaction of the expected learning goals. Second, the introduction of collaborative activities seems to have a positive effect on learning gains, thanks to the opportunity to verbalize the received information with a peer. This result supports the idea of introducing social activities in this kind of lectures.
Overall, the project was able to answer all the research questions to different extents, which makes it a successful experiment.

References

[1] *Burpees for Beginners: How to do a Burpee.* URL: `https://www.youtube.com/watch?v=dZgVxmf6jkA`.

[2] Andrea Cespedes. *What Muscles Are Used in Push-Up Exercises?* URL: `https://www.livestrong.com/article/17377-muscles-used-pushup-exercises/`.

[3] Pierre Dillenbourg. *Orchestration Graphs, Modeling scalable education.* 2015.

[4] Nicole McDermott. *How to Do the Perfect Squat.* URL: `https://greatist.com/fitness/perfect-squat`.

[5] *ShapirofffdfffdfffdWilk test.* URL: `https://en.wikipedia.org/wiki/Shapiro%5C%E2%5C%80%5C%93Wilk_test`.

[6] Jenny Sugar. *Are You Sure You're Doing Squats Correctly?* URL: `https://www.popsugar.co.uk/fitness/How-Do-Squats-41301486`.

[7] Sally Tamarkin. *A Brief History Of The Burpee.* URL: `https://www.huffpost.com/entry/burpee-history_n_5248575`.

[8] *The Perfect Push Up — Do it right!* URL: `https://www.youtube.com/watch?v=IODxDxX7oi4`.