Understanding The Sun-Earth-Moon System
PSI - IPS Experiment

Digital Education Spring 2025 Project

Mariem Baccari
Mehdi Bouguerra Ezzina
Samuel Bélisle

Ecole Polytechnique Fédérale de Lausanne
Lausanne, Switzerland

June, 2025

Contents

1 Introduction

1.1 Aims of the Study

This project investigates and compares two pedagogical strategies in designing a lesson about the Sun-Earth-Moon system:

- **The I-PS approach** is a classic one that starts by giving a lecture about the topic and continues with giving to the students exercises to train that newly acquired knowledge. We call it I-PS for "Instruction followed by Problem Solving".
- **The PS-I approach** is less common in standard education systems. It starts with giving some exercises to the students and only later giving them a lecture on the topic. We call it PS-I for "Problem Solving followed by Instruction".

Our question is: Does engaging in a problem-solving task before the formal instruction (PS-I) lead to better understanding of the lunar cycle than receiving the instruction first (I-PS) ?

Our lesson focuses on the Sun-Earth-Moon system, a domain that appears intuitive but often reveals deep misconceptions, even among scientifically literate students as confirmed by our Pre-Test results. We selected it because it combines real-world relevance with precise scientific and almost geometrical explanations.

1.2 The participants

Our target audience is EPFL students. This roughly means that they will be between eighteen and thirty years old, they will have completed a high school education and maybe even a Bachelor of Science. We then know that they will master the prerequisites. In the first hand, our prerequisites are understanding that the Earth is roughly spherical, that its rotation causes day and night and that the Moon orbits the Earth. In the other hand, the students should have a basic understanding of how a light-point will illuminate a sphere (that it will always illuminate half of it) even if they have not yet realized that this happens to the Moon as well as to the Earth.

1.3 Our Model

We use a simplified model of the Sun-Earth-Moon system that allows predictions. It adopts a fixed top-down view, perpendicular to the ecliptic plan where the Sun is a distant light source illuminating half of the Moon and Earth constantly. The Moon's orbit is modeled as circular and coplanar with the ecliptic, and Earth's axial tilt is ignored. The model allows switching between this spacial point of view to the point of view of someone on Earth. In particular, we consider this model fully acquired when the students can successfully shift from one representation to the other. The model can be explored using this simulator (link) which has been extremely helpful both as a pedagogical tool and for making sure the questions are accurate.
Limitations: The model omits the Moon's 5.14° orbital tilt, the ellipticity of its orbit, Earth's axial tilt, and the distinction between tropical and synodic months. These simplifications reduce complexity without impairing core learning objectives and are discussed more in Section 8.1.

2 Learning Goals

We do not evaluate the ability to remember terminology. Hence the cheat-sheet in figure 1 is available for the participants throughout the entire experiment. It is also especially useful for non native english speakers.

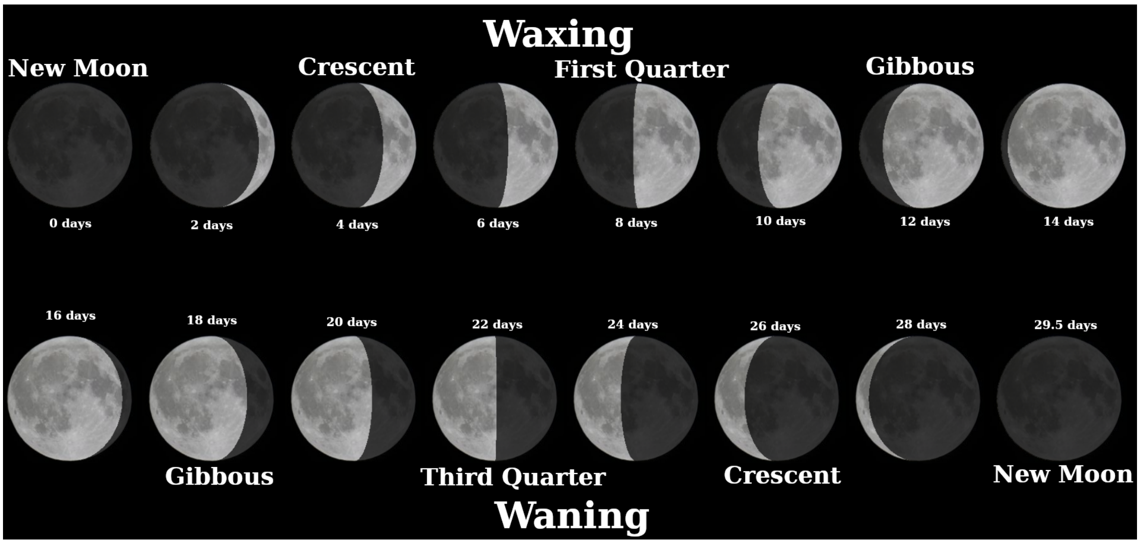


Figure 1: Cheat-sheet available throughout the study

The general learning goal is for the students to master the model presented in section 1.3. More precisely, students should be able to:

1. Predict how a ball will be enlightened based on the position of the light and infer the light direction from the ball lightning.
2. Describe the different phases of the lunar cycle.
3. Explain how the Sun-Earth-Moon system evolves during one hour, one day, one month.
4. Infer the status on the Sun-Earth-Moon system based on partial information on the system.
5. Predict the evolution of some factor (e.g. phase, sky position) after one hour, one day, one month.

We list here goal (1) for convenience but we consider it more as a prerequisite because it is quite intuitive. Our lesson do not teach it specifically but one can see it in action in the lessons (the lesson 1 especially).

Goal (2) and (3) are abilities that can be learned by heart but are linked. Knowing one lets us re-discover the other if we have a good understanding of the model. For this reason, we provided the cheat-sheet to make sure every participant has access to the phases (2) and we ask them in the PS activity about the (3) which force them to built their own model (for PS-I) or train the one learned in prior Instruction (for I-PS).

The capacity to infer (4) is essentially linked to the lighting of a ball (1) but applied to the Moon and has a lot of extrinsic cognitive load because we perceive the Moon as flat, because the shaded part has almost no light and because we do not often intuitively identify the sun as the main lighting source. Such inference tasks can be completed by shifting between the point of view of a human on Earth and a visualization of the system from space.

Finally, the prediction goal (5) necessitates a deep understanding of the cycle (2), the ball

(1), and the system (3).

We can test the their inference ability (4) together with their prediction abilities (5) by presenting partial information about the system to the students and asking them the evolution of this information or another.

3 Lesson Design and Activities

3.1 Pre- and Post-Tests

To measure learning outcomes, we use pre- and post-tests composed of multiple-choice questions aligned with our learning goals. While the tests differ in exact items, they are similar in structure and difficulty. Because the Pre-Test has seven questions and the Post-Test have eight, we give a score for each as the percentage of the maximum grade for that test. We define the learning gain for a participant as the difference between their Post-Test and Pre-Test score percentages. We discuss this choice in Section 8.3.

3.2 Problem Solving

All participants complete the same problem-solving (PS) task involving a story with missing information and related questions. Students must infer unknowns and justify their reasoning or identify the information needed to answer. This activity emphasizes metacognition by encouraging reflection on knowledge gaps and reasoning strategies in the justification box.
Previous research [1] encouraged us to pursue a story-based core activity to maintain student engagement. They need to understand the situation and answer with information they deduce from the context. They can connect with our heroes.
The deterministic nature of the system requires a single correct solution if sufficient data is provided. Questions vary in difficulty:

- Some are solvable intuitively (even without instruction).
- Most require concepts taught in the instruction.
- A few are unsolvable because we do not give key information but a student who followed the instruction should be able to eliminate some states and/or precisely pinpoint what information is missing and explain how to get the full answer from it.

This design supports different functions depending on the instructional order. In the I+PS group, the activity reinforces learned theory. In the PS+I group, it reveals knowledge gaps, potentially increasing motivation and curiosity, and force them to reevaluate their preconceptions, especially the false ones, regarding the moon cycle.

3.3 Pilots

Before starting the experiment, We conducted three pilot sessions on only the PS task:

- **Pilot 1:** The participant was already well-informed about the moon and performed very well, solving most of the questions easily. This initially suggested that the PS part could be "too easy". However, the next pilots confirmed that this was the exception and not the rule.
- **Pilot 2:** The participant struggled after relying on prior false beliefs and expressed doubts in the free space, highlighting the activity's potential to trigger both reflection and frustration. It was unclear which sentiment was prevalent here.

- **Pilot 3:** Similar difficulties were encountered where the participant expressed frustration at the task. At around ¾ of the time limit, they seemed to realize that deducing the moon's shape based on the sun's position in the sky was possible but expressed that they did not know how to do that, demonstrating a clear desire to know more.

Feedback prompted revisions to clarify that guessing or stating assumptions is acceptable. We added this note to the instructions: "You can write down an assumption if you feel we did not provide enough information."

3.4 Instruction Phase

The instruction phase consists in three lessons:

3.4.1 Lesson 1: A small demonstration experiment

This lesson introduces our students to how light and shadows work using a simple physical model that represents the system consisting in the moon, the Earth and the Sun. The students observe how light interacts with a spherical object and develop an intuitive understanding of how the percentage of illumination depends on the relative position of the object with respect to the light source and the observant. With this lesson, we aim prepare the student's cognitive state for the formal description of the model and make sure they have the learning goal 1 and apply it to the Moon.

3.4.2 Lesson 2: Basic concepts around the moon cycle

This lesson presents the dynamics of the Sun-Earth-Moon system by comparing the point of view from space and from the earth. It repeats the explanation about the orbit and the phases but now illustrates it with realistic representations of the sky and uses the 4-weeks timeline to completely address the learning goal 2.

3.4.3 Lesson 3: Evolution of the lunar cycle via simulation

This lesson aims to deepen the understanding of the previous concepts by focusing on how the lunar cycle evolves over time. We observe the Sun-Earth-Moon system with different time granularities (hours / days) and start by going over one day hour by hour during a waning gibbous phase then one month day by day. Our aim is to consolidate the knowledge from a time perspective and reach the learning goal 3.

These lessons are based on switching between multiple perspective to enhance conceptual flexibility and help the participants apply the knowledge across various contexts. We provide direct instruction and exploratory activities that follow a scaffolding approach by having an explicit instruction in lesson 2 before moving to guided discovery and self-directed exploration in lesson 3, supporting a gradual release of responsibility. The lesson 1 makes sure that the geometrical prerequisite is understood by the student.

4 Experimental Design

As a reminder, our research question is: **Does engaging in a problem-solving task before the formal instruction (PS-I) lead to better understanding of the lunar cycle than receiving the instruction first (I-PS) ?**

To answer this, we split our subjects into two groups:

- **A Control Group (I-PS)** which was instructed to complete the instruction first, followed by the problem-solving task.
- **A Test Group (PS-I)** which was instructed to complete the problem-solving task first, followed by the instruction.

They all complete a pre-test and a post-test, respectively before and after the assigned sequence of tasks. The pre-test scores allows us to determine the initial knowledge and the post-test to measure the overall learning outcomes.

4.1 Variables

The variables collected in this study can be categorized into two distinct groups: demographic variables that provide insight into participant backgrounds and characteristics, and experimental variables that measure performance outcomes and engagement behaviors during the study.

4.1.1 Demographics variables

Demographic data were collected to characterize the participant sample. These variables included academic program (Bachelor's, Master's, etc.), field of study (section), age, gender, and primary geographical region where participants had lived most of their lives.
The collection of demographic information served dual purposes. First, it enabled a comprehensive understanding of participant backgrounds, including potential prior knowledge due to academic experiences, cultural factors that might influence performance on the tests or introduce systematic biases. This contextual information is essential for interpreting results within the specific population studied. Second, demographic data allowed for balancing similarly the I-PS and PS-I group, helping to minimize confounding variables that could contaminate the comparison of instructional methods.

4.1.2 Experiment-related variables

The primary experimental measures focused on learning outcomes and engagement patterns. Pre-test and post-test scores were collected from all participants to establish baseline knowledge levels and measure learning gains following the instructional interventions. These scores form the foundation for comparing the relative effectiveness of the PS-I and I-PS approaches.
Additionally, for participants in the PS-I group, time spent completing the problem-solving component was recorded as a measure of active engagement. This timing variable serves as a behavioral indicator of participant involvement and allows for investigation of the relationship between engagement duration and learning outcomes. Specifically, it enables testing of the hypothesis that participants who invest more time in problem-solving activities demonstrate greater improvement in post-test performance.

4.2 Motivation

4.2.1 I-PS (Control Group)

We expect the problem solving part to reinforce learning by consolidating concepts introduced during the instruction phase. One limitation could be that the instruction overwhelms the students with too many unfamiliar concepts that they are not yet prepared to grasp. As engaging in actual reasoning is critical for the student's development, we prohibited using the simulator during problem solving.

4.2.2 PS-I (Experimental Group)

We hypothesize that this format will give the student a desire for knowledge. If this works, the student in the Test Group should follow the Instruction phase better. This is the "Productive failure" effect [1]. Additionally, we expect that the PS activity acts as an "Advanced Organiser" [2] which activates some of their cognitive processes that will be mobilized during Instruction and helps them make links with other concepts (in particular from their every-day life). If this works, the students will have a more profound understanding of the concepts taught during instruction.

5 Implementation

The full lesson was implemented as two online Moodle courses called IPS/PSI + "Understanding The Sun-Earth-Moon System - An Experiment" to ensure a controlled and trackable learning environment that allows participants to complete the activities asynchronously and at their own pace which was a crucial element for us to be able to recruit enough participants. The I-PS and PS-I are copies with only the order of the instruction phase and the problem solving activity reversed. The I-PS course can be found here (link) and the PS-I course here (link).
The course consisted in a sequential series of the following elements:

5.1 Information on the Study

Here we simply give an overview of the context of this experiment and the rights of the participants. We also emphasize that the study is not anonymous (we sometimes had to help them with technical issues during the experiment and used their username to join the data and check completion) but that the data will only be accessed by people involved in this course and deleted at the end.

5.2 Basic Information

A series of questions in quiz form to learn more about our demographic. More specifically, we asked them to confirm that they are registered in EPFL, the name of the program, section, their age, gender and the primary geographical region where they have lived. This last question let us know from which latitude their are used seeing the Moon, which has a huge impact on the observations.

5.3 Pre-Test and Post-Test

As demonstrated in figures 2 and 3, both tests were a multiple choice quiz with fixed answers. Depending on the question, one or many answers could apply. We provided a cheat-sheet with the lunar cycle terminology throughout the experiment.

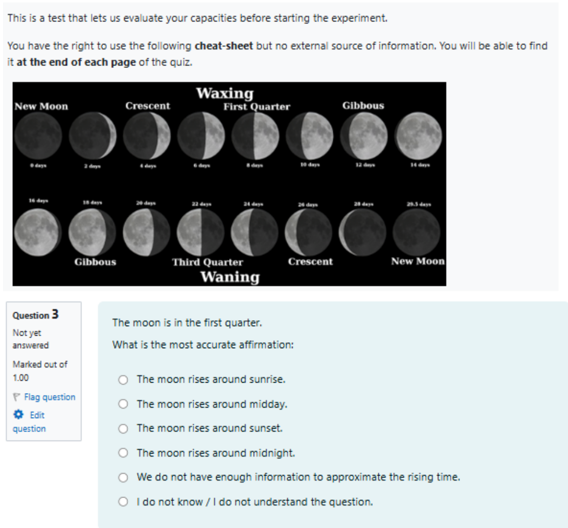


Figure 2: Screenshot of the presentation for the pre-test quiz and of one sample question.

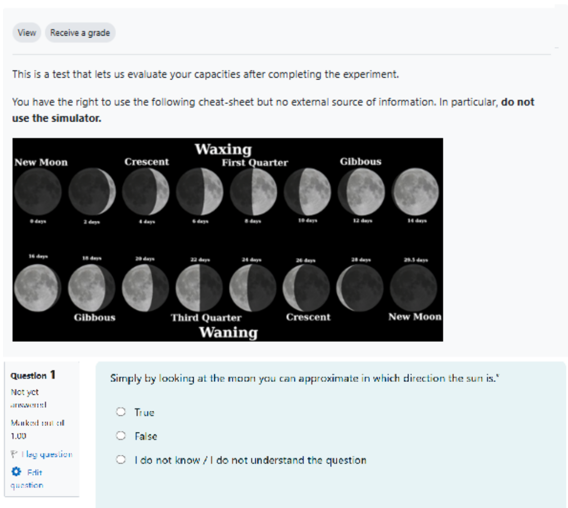


Figure 3: Screenshot of the presentation for the post-test quiz and of one sample question.

5.4 Lesson

A series of three video lessons directly embedded into Moodle.

- **Lesson 1:** A short video by the National Science Teaching Association after which students are encourage to replicate the experiment using a flashlight and a ball. This task was, of course, optional due to the asynchronous nature of our course but framed as beneficial for the students' understanding.
- **Lesson 2:** An animated video by timeanddate.
- **Lesson 3:** A sequence of screen-recorded videos by a team member as shown in figure 6 demonstrates how to use an open-source web-based Lunar Phase Simulator (link) as a software then as part of two structured scenarios.

Figures 4 and 5 show, each, two screenshots from the the lessons 1 and 2 that demonstrate the switch in perspective discussed in the lesson design section.

Figure 4: Screenshots showing the switch in perspective in the lesson 1 video.

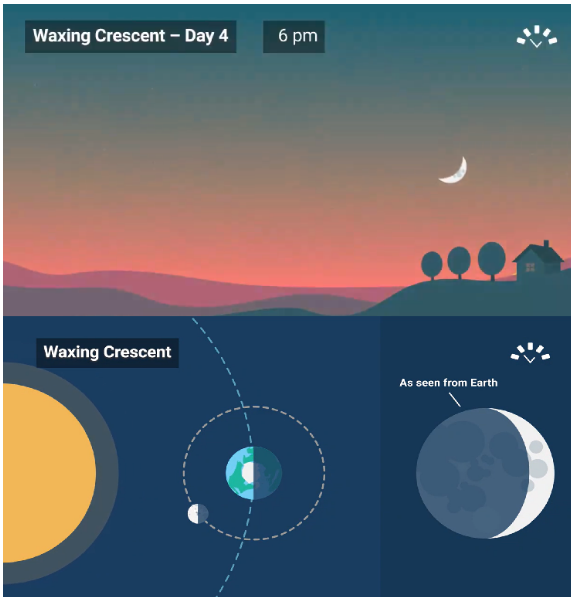


Figure 5: Screenshots showing the switch in perspective in the lesson 2 video.

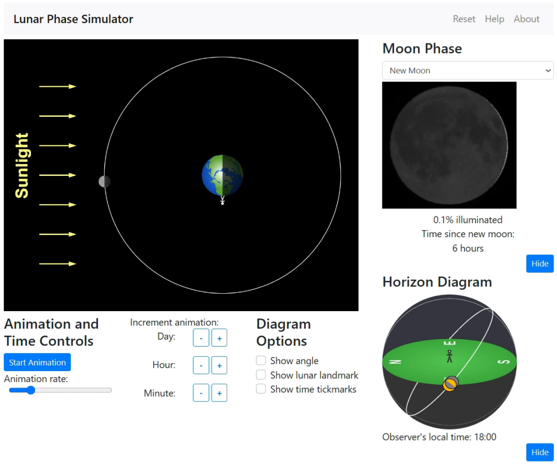


Figure 6: Screenshot from lesson 3.

To maintain student engagement, we kept the videos short such that the total length of the instruction phase is around 10 minutes.

5.5 Problem Solving Activity

This was implemented as a quiz with a time limit of 20 minutes. The activity was story-based with no pre-defined answers. The student could answer in text and we always included in the question "How do you know or what information would you need in order to answer precisely?" to encourage them to fully express their reasoning.

5.6 Usability

Considering the nature of our participants' demographic, the Moodle interface was intuitive for all of them although the account creation and some initial technical issues required support and continuous improvements to the course implementation.

5.7 Group Management

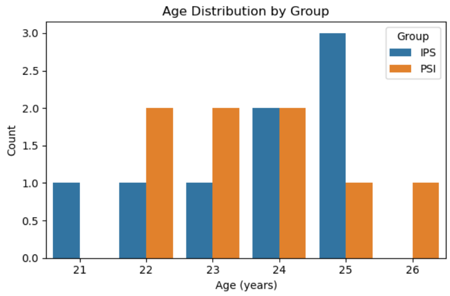
We assigned participants to the I-PS or PS-I groups at first randomly and later on, keeping an eye on the numbers and demographics to keep them balanced.

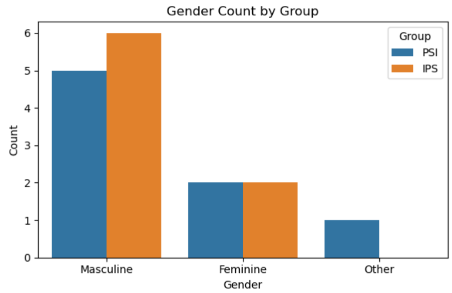
6 Participants, Data, and Analysis

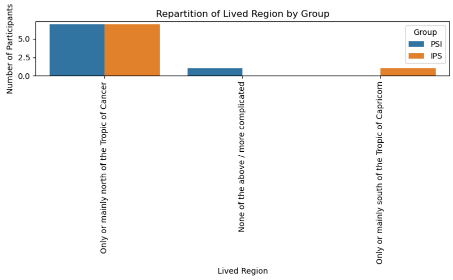
This section details the characteristics of the participants and the statistical methods employed to analyze the data and compare the effectiveness of the PS-I and I-PS teaching approaches.
The data and a jupiter notebook are available in this public GitHub repository (link).

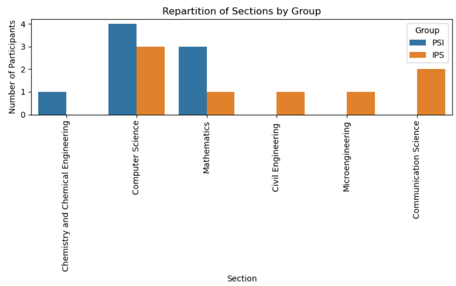
6.1 Participants

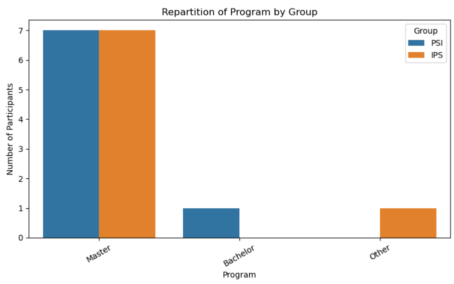
A total of 16 students from the École Polytechnique Fédérale de Lausanne (EPFL) participated in this study. Participants were equally distributed between two experimental conditions. The age of participants ranged from 21 to 26 years. There were 14 master's students. A majority of participants were enrolled in Mathematics or Computer / Communication Science sections. Gender distribution was balanced across groups with 2 women in each condition, which is somewhat representative of the gender repartition of the school (around 30% of the students). Finally most of the participants reported having lived north of the Tropic of Cancer most of their life. The following plots shows these demographics information.











6.2 Data analysis

Each participant completed a pre-test assessing baseline knowledge, engaged in the lesson according to their assigned condition, and then took a post-test at the end of the lesson. For the PS-I group only, we take into consideration the time spent on the initial problem-solving activity.

6.2.1 Descriptive statistics

From pre- and post-test scores difference, we computed a Learning Gain score. The Table 1 summarizes mean scores, standard deviation and the median by group.

Group	PreTest_grade				PostTest_grade				Learning_Gain			
	count	mean	std	median	count	mean	std	median	count	mean	std	median
I-PS	8	41.07	16.53	42.86	8	53.13	23.86	50.0	8	12.05	20.53	12.50
PS-I	8	48.21	32.11	42.86	8	56.25	16.37	56.25	8	8.04	27.16	7.15

Table 1: Details about percentages of correctness for both groups.

Both groups have an increased score from pre-test to post-test, with mean learning gains of 12.05% for the I-PS group and 8.04% for the PS-I group. Notably, while the PS-I group started with a higher pre-test mean (48.21% vs. 41.07%), both groups achieved similar post-test performance levels (56.25% vs. 53.13%). We show an interaction graph with the tests averages and boxplots of the repartition in Figure 7. The lines are near from parallel and the noise is big which shows visually that the effect is really small and not significant.

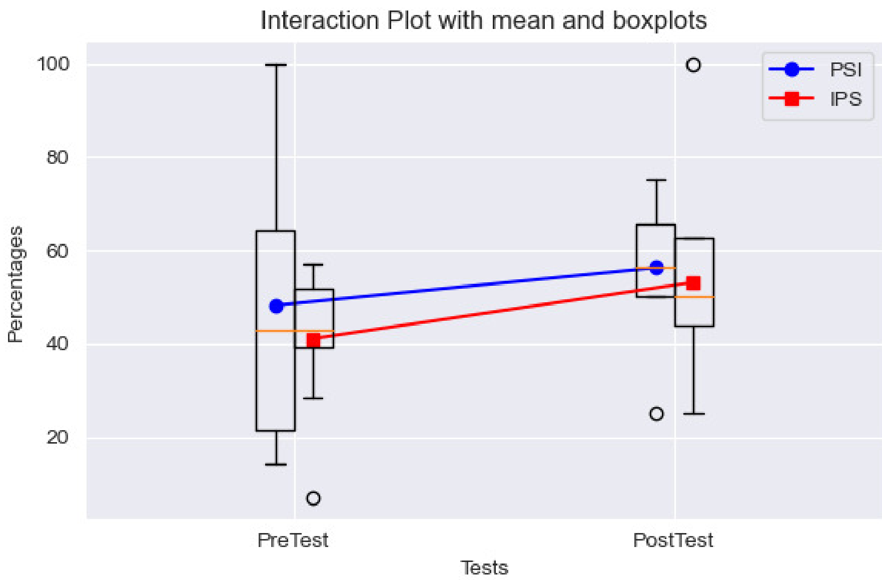
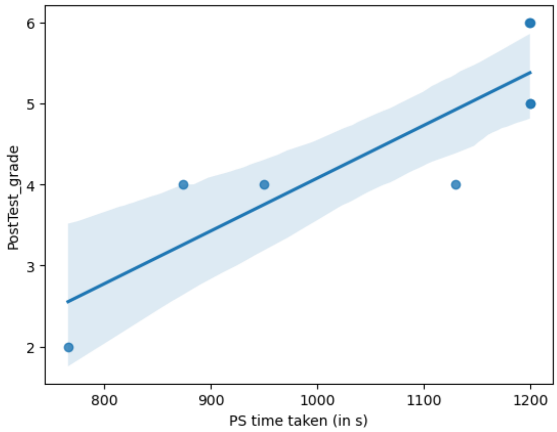


Figure 7: Interaction graph showing parallels lines.

6.2.2 Exploratory Analysis

Correlation

Pearson correlation analysis revealed a very strong positive linear relationship ($r = 0.874$, $p = 0.005$) between time spent on the problem-solving part and post-test performance for the PS-I subjects. This statistically significant correlation suggests that participants who spent more time engaged in the activities achieved higher scores.



Group Comparison
To test the primary hypothesis regarding the effectiveness of PS-I versus I-PS approaches, both one-way ANOVA and independent samples t-test were conducted on learning gains. Both tests yielded non-significant results for the p-value ($p = 0.744 > 0.05$), indicating no statistically significant difference in learning gains between the two instructional approaches.

Effect Size and Power Analysis
We calculated Cohen's d for the difference in learning gain (PS-I vs. I-PS): $d = -0.167$. On average, the PS-I group improved slightly less than the I-PS group, but the effect seems negligible according to Cohen's benchmark. The two methods produced almost identical mean gains with our participants.
Nevertheless, an important critic that can be made of all our previous observation is that they have been made on a very small number of participants. Effectivelly the p-value heavily depends on the sample size and, for small one's, Cohen's d can overestimate the "true" population effect size.
That is why we have done a post-hoc power analysis (two-tailed, $\alpha = 0.05$) based on $|d| = 0.167$ which revealed an achieved power of only 0.061. To attain 80% power for detecting an effect of this size, a total sample size of approximately 1,130 participants would be required. This highlights that the present study is substantially underpowered to detect small effects.

7 Conclusions and Reflections

The most critical limitation was the severely underpowered design (power = 0.058), requiring approximately 1,572 participants to detect meaningful effects. Future studies should conduct a priori power analyses to determine adequate sample sizes before data collection begins.

We should add more control variables notably to be aware of pre-existing domain knowledge (beyond conducting a simple pre-test).

Conducting delayed post-tests (1 week, 1 month, 3 months) to examine whether PS-I vs I-PS differences emerge over time. Some research suggests that productive failure effects may manifest in long-term retention rather than immediate performance.

8 Annexes

8.1 Detailed limitations of our model

As mentioned in the introduction, the model we teach have a few limitations that we had not the time to cover in the lesson. In particular, the orbit of the Moon is not within the ecliptic plane but has a 5.14° orbital inclination (link). This is the reason why solar and lunar eclipses are rare and not occurring every two weeks. Moreover, the real orbit of the Moon around Earth is not a circle but an ellipse with an eccentricity of around 5% which makes it size change.
These two approximation have an impact on the observations, but we did not teach about them (nor tested knowledge about them).

An important aspect of Earth's orbit that is entirely neglected in this model is its orbit's inclination which causes the seasons and the duration of day and night to vary during the year. Gladly, this inclination have only a limited impact on the moon and we took necessary measures to avoid them conflicting with our experiment, such as placing our Problem Solving exercise in February.

On a more technical note, our explanation completely glossed over the distinction between tropical month (27.3 days which is the duration of the Moon's orbit **relative to the stars**) and synodic months (29.5 days which is the duration of the Moon's orbit **relative to the sun**, used in our model), which can lead to confusion when hearing about any of those duration and thinking about the other, but when considering phases of the moon, tropical months are irrelevant.

8.2 Various qualitative observations

We here present remarks made by us or received from the participants. The participants are presented anonymously using their id from the data file which reflect in which group they were.

PSI1 commented about the time-limit on the PS activity. They commented that the activity was too hard/long to be completed in that time. This was expected as they were in the PS-I group. Note that all participants were warned that the question were difficult and that we did not expect high accuracy nor completion.

IPS1 made some drawings during the Pre-Test which allowed them to re-construct a model really similar to the one presented afterwards. However, they were missing key elements and expressed relief while receiving them in the instruction. They managed a perfect score at the Post-Test, using their fixed way of drawing the situation.

IPS8 had a technical issue just after the Pre-Test, blocking them from continuing immediately into instruction. While exchanging about how to continue, they said that the Pre-Test has *hyped* them. More precisely, they noticed during the Pre-Test that the knew less than they taught and this made them look forward to continue the activity. This is noticeable, because it describes precisely one of the feelings aimed by the problem solving in the PS-I condition, and goes in the direction of the Productive Failure effect. Ironically, this participant was not in PS-I but in I-PS and received this effect with the Pre-Test only, hinting that our Problem Solving's effect might be less big in this experiment situation compared with an instruction context, without a Pre-Test. This situation hints to another design experiment presented in Section 8.3 below.

We want to add that other participants had a strong Productive Failure effect. In particular, *PSI8* said that the problem solving part was extremely frustrating. They were wondering why the videos were not before this exercise. But they then added that this made them ultra-focus during the videos, trying to gather as much knowledge as possible to avoid a similar situation afterwards.

Similarly, during the Pilot for the Problem Solving, one of the participants said “I certainly look stupid right now” which is in indicator that the Productive Failure effect is either working well or too pronounced.

8.3 Comparison between Pre-Test and Post-Test

As said in Section 3.1, we designed the questions of the Pre and Post test such that they should be more or less equivalent. However, we can’t be sure that they truly were and qualitative observation seem to indicate that they were probably not. In particular, one of the PS-I participants obtained the full seven points on the Pre-Test, managed the problem solving activity very well and then, after the instruction, achieved five out of the eight points of the Post-Test. This hints that the questions were probably trickier in the Post-Test.

However, this changes only slightly the interpretation of the learning gain (negative score is not necessarily linked to regression), but does not impact the comparison between PS-I and I-PS.

If we had more participants we would have liked to test another condition where participants do only the Pre- and Post-Tests without any instruction nor problem solving in between. This would help with assessing the apriori knowledge of our population and with comparing the Pre- and Post-Tests.
Another way to solve the issue would have been to change the experimental framework. With a lot of participants, we could have three groups. A control group which only does a test compared with PS-I and I-PS groups which start with their lesson and finish with this same test. The learning gain would then be only the difference between the result at the test compared with the average result in the population. This approach would not have the opportunity of looking at the specific learning gain of one participants but would have a *improvement of score compared with projected prio results.*

References

[1] Manu Kapur and Katerine Bielaczyc. Designing for productive failure. *Journal of the Learning Sciences*, 21(1):45–83, 2012.

[2] David P. Ausubel. The use of advance organizers in the learning and retention of meaningful verbal material. *Journal of Educational Psychology*, 51(5):267–272, 1960. `doi:10.1037/h0046669`.