Journal of Visualized Experiments

Problem-Solving Before Instruction (PS-I): A protocol for assessment and intervention in students with different abilities

--Manuscript Draft--

Article Type:	Invited Methods Collection - Author Produced Video
Manuscript Number:	JoVE62138R4
Full Title:	Problem-Solving Before Instruction (PS-I): A protocol for assessment and intervention in students with different abilities
Corresponding Author:	Eduardo González-Cabañes University of Oviedo: Universidad de Oviedo Oviedo, Asturias SPAIN
Corresponding Author's Institution:	University of Oviedo: Universidad de Oviedo
Corresponding Author E-Mail:	gonzalezventura@uniovi.es
Order of Authors:	Eduardo González-Cabañes
	Trinidad García
	José Carlos Núñez
	Celestino Rodríguez
Additional Information:	
Question	**Response**
Please indicate whether this article will be Standard Access or Open Access.	Standard Access (US$1200)
Please confirm that you have read and agree to the terms and conditions of the author license agreement that applies below:	I agree to the Author License Agreement
Please specify the section of the submitted manuscript.	Behavior
Please indicate whether this article will be Standard Access or Open Access.	Standard Access ($1400)
Please confirm that you have read and agree to the terms and conditions of the video release that applies below:	I agree to the Video Release
Please provide any comments to the journal here.	

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

Manuscript Click here to access/download;Manuscript;62138_R4.docx

TITLE:

Problem-Solving Before Instruction (PS-I): A Protocol for Assessment and Intervention in Students with Different Abilities

AUTHORS:

Eduardo González-Cabañes[1]*, Trinidad García[1]*, José Carlos Núñez[1]*, Celestino Rodríguez[1]*

[1]Department of Psychology, Oviedo University, Spain

*These authors contributed equally.

Email addresses of co-authors:

Trinidad García (garciatrinidad@uniovi.es)
José Carlos Núñez (jcarlosn@uniovi.es)
Celestino Rodríguez (rodriguezcelestino@uniovi.es)

Corresponding author:

Eduardo González Cabañes (gonzalezventura@uniovi.es)

KEYWORDS:

Invention Activities, Productive Failure, Exploratory Learning, Self-regulated Preparation for Learning, Self-efficacy, Mastery Approach Goals, Metacognition, Divergent Thinking, Curiosity, Transfer, Cognitive Load, Statistics Education.

SUMMARY:

This protocol guides researchers and educators through implementation of the Problem-Solving before Instruction approach (PS-I) in an undergraduate statistics class. It also describes an embedded experimental evaluation of this implementation, where the efficacy of PS-I is measured in terms of learning and motivation in students with different cognitive and affective predispositions.

ABSTRACT:

Nowadays, how to encourage students' reflective thinking is one of the main concerns for teachers at various educational levels. Many students have difficulties when facing tasks that involve high levels of reflection, such as on STEM (Science, Technology, Engineering and Mathematics) courses. Many also have deep-rooted anxiety and demotivation towards such courses. In order to overcome these cognitive and affective challenges, researchers have suggested the use of "Problem-Solving before Instruction" (PS-I) approaches. PS-I consists of giving students the opportunity to generate individual solutions to problems that are later solved in class. These solutions are compared with the canonical solution in the following phase of instruction, together with the presentation of the lesson content. It has been suggested that with this approach students can increase their conceptual understanding, transfer their learning to different tasks and contexts, become more aware of the gaps in their knowledge, and generate a personal construct of previous knowledge that can help maintain their motivation. Despite the

advantages, this approach has been criticized, as students might spend a lot of time on aimless trial and error during the initial phase of solution generation or they may even feel frustrated in this process, which might be detrimental to future learning. More importantly, there is little research about how pre-existing student characteristics can help them to benefit (or not) from this approach. The aim of the current study is to present the design and implementation of the PS-I approach applied to statistics learning in undergraduate students, as well as a methodological approach used to evaluate its efficacy considering students' pre-existing differences.

INTRODUCTION:

One of the questions that teachers are most concerned about currently is how to stimulate students' reflection. This concern is common in courses of a mathematical nature, such as STEM courses (Science, Technology, Engineering and Mathematics), in which the abstraction of many concepts requires a high degree of reflection, yet many students report approaching these courses purely through memory-based methods[1]. In addition, students often show superficial learning of the concepts[1-3]. The difficulties that students experience applying reflection and deep learning processes, however, are not only cognitive. Many students feel anxiety and demotivation faced with these courses[4,5]. In fact, these difficulties tend to persist throughout students' educations[6]. It is therefore important to explore educational strategies that motivationally and cognitively prepare students for deep learning, regardless of their differing predispositions.

It is particularly useful to find strategies that complement typical instructional approaches. One of the most typical being direct instruction. Direct instruction means fully guiding students from the introduction of novel concepts with explicit information about these concepts, then following that with consolidation strategies such as problem-solving activities, feedback, discussions, or further explanations[7,8]. Direct instruction can be effective for easily transmitting content[8-10]. However, students often do not reflect on important aspects, such as how the content relates to their personal knowledge, or potential procedures that could work and do not[11]. It is therefore important to introduce complementary strategies to make students think critically.

One such strategy is the Problem-Solving before Instruction (PS-I) approach[12], also referred to as the Invention approach[11] or the Productive Failure approach[13]. PS-I is different to direct instruction in the sense that students are not directly introduced to the concepts, instead there is a problem-solving phase prior to the typical direct instruction activities in which students seek individual solutions to problems before getting any explanation about procedures for solving them.

In this initial problem, students are not expected to fully discover the target concepts[13]. Students may also feel cognitive overload[14-16] and even negative affect[17] with the uncertainty and the many aspects to consider. However, this experience can be productive in the long term because it can facilitate critical thinking about important features. Specifically, the initial problem can help students to become more aware of the gaps in their knowledge[18], activate prior knowledge related to the content to cover[13], and increase motivation because of the opportunity to base

their learning on personal knowledge[7,17,19].

In terms of learning, the effects of PS-I are generally seen when the results are evaluated with deep learning indicators[20,21]. In general no differences have been found between students who learned through PS-I and those who learned through direct instruction in terms of procedural knowledge[20,22], which refers to the ability to reproduce learned procedures. However, students who go through PS-I generally exhibit higher learning in conceptual knowledge[7,19,23], which refers to understanding the content covered, and transfer[7,15,24], which refers to capacity to apply this understanding to novel situations[7,15,19,24]. For example, a recent study in a class about statistical variability showed that students who were given the opportunity to invent their own solutions to measure statistical variability before receiving explanations about the general concepts and procedures in this topic demostrated better understanding at the end of the class than those who were able to directly study the relevant concepts and procedures before getting involved in any problem-solving activity[23]. However, some studies have shown no differences in learning[16,25,26] or motivation[19,26] between PS-I and direct instruction alternatives, or even better learning in direct instruction alternatives[14,26], and it is important to consider potential sources of variability.

The design features underlying the implementation of PS-I are an important feature[20]. A systematic review[20] found that there was more likely to be a learning advantage for PS-I over direct instruction alternatives when the PS-I interventions were implemented with at least one of two strategies, either formulating the initial problem with contrasting cases, or building the subsequent instruction with detailed feedback about the students' solutions. Contrasting cases consist of simplified examples that differ in a few important characteristics[11] (see **Figure 1** for an example), and can help students identify relevant features and evaluate their own solutions during the initial problem[11,20]. The second strategy, providing explanations that build on the students' solutions[13], consist of explaining the canonical concept while giving feedback about the affordances and limitations of solutions generated by students, which can also help students focus on relevant features and evaluate the gaps in their own knowledge[20], but after the initial problem-solving phase is completed (see **Figure 3** for an example of the scaffolding from students' typical solutions).

Given the support in the literature for these two strategies, contrasting cases and building instruction on students' solutions, it is important consider them when promoting the inclusion of PS-I in real educational practice. This is the first goal of our protocol. The protocol provides materials for a PS-I intervention that, while adaptable, is contextualized for a lesson on statistical variability, a very common lesson for university and high school students, who are generally the target populations in the literature on PS-I[29]. The initial problem-solving phase consists of inventing variability measures for income distributions in countries, which is a controversial topic[30] that may be familiar to students in many learning areas. Then materials are provided for students to study solutions to this problem in a worked example, and for a lecture that incorporates discussion of common solutions produced by students along with embedded practice problems.

The second goal of our protocol is to make the experimental evaluation of PS-I accessible to educators and researchers, which can facilitate the actualization of the PS-I literature, including this protocol. The experimental evaluation described in the protocol can be applied in ordinary lessons, since students in a single class can be assigned the materials for the PS-I condition or the materials for a direct instruction condition at the same time (**Figure 4**). This direct instruction condition is also adaptable to research and education needs, but as originally described in the protocol students start by getting the initial explanations about the target concept with the worked example, and then consolidate this knowledge with a practice problem (only presented in this condition to compensate for the time PS-I students spend on the initial problem), and with the lecture[23]. Potential adaptations include starting with the lecture and then having students to do the problem-solving activity, which is a typical control condition for comparing PS-I that has often led to better learning for the PS-I condition[7,13,19,26]. Alternatively, the control condition can be reduced to the exploration of a worked example followed by the lecture phase, which, although a more simplified version of direct instruction approaches than originally proposed, is more common in the literature and has led to varied results, with some studies indicating better learning in PS-I[15,24], and others indicating better learning from this type of direct instruction condition[14,26].

Finally, a third goal of the protocol is to provide resources for evaluating how students with different predispositions and cognitive abilities can benefit from PS-I[15]. The evaluation of these predispositions is especially important if we consider the negative predispositions that some students often have with STEM courses, and the fact that PS-I can still produce negative reactions in some cases[14]. There is, however, little research on this.

On the one hand, since PS-I facilitates the association of learning with individual ideas, rather than just formal knowledge, PS-I can be hypothesized as being able to help motivate students from low academic levels, those who have low feelings of competence, or low motivation about the subject [13,27]. One study showed that students with low mastery orientation, i.e., fewer goals related to personal learning, benefited more from PS-I than those with higher motivation to learn[27]. On the other hand, students with other profiles might encounter difficulties when involved in PS-I. More specifically, metacognition plays an important role in PS-I[31], and students with low metacognition skills might not benefit from PS-I due to difficulties in being aware of their knowledge gaps or discerning relevant content[15]. In addition, as the initial phase of PS-I is based on the production of individual solutions, students with low divergent abilities, difficulties generating a variety of responses in a given situation, might benefit less from PS-I than other students. The protocol presents reliable instruments to assess for these predispositions (**Table 1**) although others may be considered.

In summary, this protocol aims to make an implementation of a PS-I intervention that follows accepted principles in the PS-I literature accessible to educators and researchers. Additionally, the protocols provide an experimental evaluation of this intervention, and facilitate the evaluation of students' cognitive and motivational predispositions. It is a protocol that does not require access to new technologies or specific resources, and one that can be modified based on research and educational needs.

PROTOCOL:

This protocol follows the Helsinki Declaration of Ethical Principles for Research with Humans, but applies these principles to the added difficulties of integrating research within real-life settings in education[32]. Specifically, neither the assignment of learning conditions nor the decision to participate can have consequences for students' learning opportunities. In addition, confidentiality and the anonymity of students is maintained even when it is the teachers who are in charge of the evaluation. The aims, scope, and procedures of the protocol have been approved by the Research Ethics Committee of the Principality of Asturias (Spain) (Reference: 242/19).

Please note that if the user is only interested in implementing the PS-I approach, only Step 6 (without assigning participants to the control condition) and Step 7 are relevant. Despite that, Steps 5 and 9 can be added as practice exercises for students. If the user is also interested in the experimental evaluation, it is important that students work individually during Steps 4, 5, 6, and 9. It is therefore recommended that during these steps, student seating is arranged so that there is an empty space beside each student.

Depending on convenience, the steps can be implemented continuously within a single class session or with subsequent steps in a different class session.

1. Information for students about the purpose and procedures of the study

1.1. Take 10 minutes of a class period to inform students about the study.

1.2. Explicitly explain to students the general purpose of the study, their freedom to consent to participate, the fact that they may freely withdraw, and the assurance of anonymity and confidentiality in the data processing.

1.2.1. Tell them that the general purpose of the study is to explore the efficacy of different educational approaches, as well as to evaluate the influence of the students' cognitive and affective dispositions on the efficacy of these approaches.

1.2.2. Tell them that although they will be assigned to one of the two approaches, the content covered in the two conditions will be the same. Inform them that the activities used in both conditions will be available to all students at the end of the study.

1.2.3. Let them know that they are free to participate in the study and that they can leave the study at any time without affecting their learning opportunities or their grades. If they do not want to participate in the study, they can do the learning activities without handing them in. In addition, during the short time participants are completing questionnaires, non-participants can study other materials.

1.2.4. Inform them that their participation will be anonymous and that confidentiality will be

maintained at all times, an arbitrary identification number will be used to combine the data across different sessions and activities.

1.3. Provide students with two copies of the informed consent form (**Appendix A**) which also contains the researcher's contact information. Ask them to sign one copy for you, and to keep the other copy for themselves.

NOTE: This protocol is aimed at university students, where no parental permission is needed. It could be generalized to lower educational levels, although for students who are legally minors, parental informed consent would also be needed.

1.4. If students are added to the study in later phases of the protocol, ask them to complete the informed consent as described in this section before they join the study.

2. Providing students with an identification number disassociated from other records

2.1. To maintain the anonymity of students' responses, randomly assign each student an identification number (e.g., prepare a bag with random numbers and ask each student to pick one, email each student a random number through a web application). Ask them to note the number in a place where it will be accessible in the subsequent evaluations in the protocol.

NOTE: If the study is done through an online application that allows student responses to be anonymously tracked, this is not necessary.

3. Completion of questionnaires about cognitive and affective predispositions and basic demographic data

3.1. Reserve 10 minutes in a class period to administer the questionnaires to all students in the class.

3.2. Give the students who decide not to participate in the experiment other learning options such as working individually on other content.

3.3. Ask students to complete the questionnaires about their predispositions, this may be done using the questionnaires in **Appendix B**. Ask them to work individually.

NOTE: The set of questionnaires in **Appendix B** includes the Cognitive Competence Scale in the Survey of Attitudes towards Statistics (SATS-28) [33], the Mastery Approach Scale in the Achievement Goal Questionnaire-Revised[34] , the Regulation of Cognition Scale of the Metacognitive Awareness Inventory[35], and demographic questions.

3.3.1. To control for potential contaminant effects related to the order in which students complete the questionnaires, randomly hand different versions of the questionnaire sheets that vary in the order in which the questionnaires are presented. In **Appendix B-1** there are different

printed versions of the proposed questionnaires with different orders.

NOTE: If the questionnaires are completed digitally, create links with the different orders, and randomly distribute the four links among the students in the class (e.g., across groups created by alphabetic order).

3.4. Give students 7 minutes to complete the questionnaires. Instructions are included in the questionnaires and no additional instructions are needed.

4. Administration of the divergent thinking test

4.1. In case this test is of interest, take 10 minutes in a class period to administer the Alternative Uses Task[36,37] which measures fluency of divergent thinking for all students in the class.

4.2. Provide each student with blank paper and ask them to write their identification number.

4.3. Explain the instructions of the test.

4.3.1. Tell them that they will be provided with an object that has a common use, but they should come up with as many other uses as they can.

4.3.2. Give them an example (e.g., for instance, if I present you with a newspaper, which is commonly used to read, you have to write alternative uses, such as using it as a temporary hat to protect you from the sun, or to line the bottom of a travel-bag)[38].

4.4. Read the first item in the test aloud, and write it on the blackboard: "Write as many uses you can think of for a brick". Give students two minutes to write their responses. Once the two minutes are over, ask students to flip their paper to the other side.

4.5. Read the second item in the test aloud, and write it on the blackboard: "Write as many uses you can think of for a paper clip". Give students two minutes to write their responses.

4.6. Once the two minutes are over, ask the students to stop writing, and collect their papers.

5. Completion of the pre-test of previous academic knowledge

5.1. Reserve 15 minutes in a class period to administer the previous academic knowledge pre-test in **Appendix C**.

NOTE: The pre-test is about central tendency, which is relevant in order to assimilate the content on variability to be learned in the subsequent learning conditions in Step 6[7]. No class content about central tendency should be given to students between the administration of this pre-test and Step 6. We also do not recommend substituting this pre-test with a different pre-test covering variability because that can create a PS-I effect that may contaminate the results of the

experiment [26].

5.2. Distribute the pre-test to the students. From this point, ask them to work individually.

5.2.1. Give students 10 minutes to complete the pre-test. Instructions are included in the test and no more specifications are needed. Once the time is up ask the students to flip their paper over and hand it in to you.

6. Assignment to and administration of the two learning conditions

6.1. Take 35 minutes of a class period to administer the two learning conditions within the same classroom.

NOTE: To prevent reliability errors due to time, we recommend no more than one week between the completion of the questionnaires and tests in Steps 2 and 3 and this step.

6.2. Ensure that the task books are properly prepared, containing the materials for the two conditions.

NOTE: GDP per capita has been chosen to contextualize these learning materials for several reasons: firstly, it is a controversial topic[30] that may be familiar to students from many learning areas, and secondly it is a ratio variable that allows the use of different variability measures that are discussed during the lesson (range, interquartile range, standard deviation, variance, and coefficient of variation).

6.2.1. For the PS-I condition, print the corresponding task book in **Appendix D-1** which contains: the Invention Problem activity, in which students are asked to invent an inequality index; the Worked Example activity, in which students can study the solutions for this problem.

6.2.2. For the direct instruction condition, print the corresponding task book in **Appendix D-1** which contains: the Worked Example activity (the same Worked Example given to the PS-I condition); the Practice Problem paired with this Worked Example.

NOTE: It is important that the practice problem included in the materials for this condition is not present in the PS-I condition. It is included to experimentally compensate for the extra time spent by the PS-I students on the invention problem. An intrinsic limitation of PS-I designs is the difficulty to control for equivalence in terms of both time and materials. Even in designs in which the PS-I condition and the control condition only differ in the order in which learning materials are presented (that is, either presenting a problem *before* an explicit instruction phase, or presenting the exact same problem *after* the exact same explicit instruction phase), equivalence is not achieved, because a problem that is solved before instruction is expected to take more time than after instruction. This protocol deals with this problem in the same way as other studies[24], by including extra materials in the direct instruction condition.

6.2.3. Separate the two activities in each task book by binding the papers corresponding to the second activity (e.g., with a clip or a sticky note) together so that students cannot see the contents of the second activity while they are doing the first activity.

6.3. Inform students of the procedure to follow in this specific step.

6.3.1. Tell them that depending on the task book they are assigned, they will have two different pairs of activities, but all students will see the same content, and at the end of the lesson all of them will have access to all of the activities.

6.3.2. Let them know that they will be told when to start the first activity and when they should move to the second activity. Also tell them that the papers for the second activity have been bound to prevent them from looking before the appropriate time.

6.3.3. To reduce potential frustration related to fear of failing, tell them that although they might find some activities difficult, they should try to see these difficulties as learning opportunities[39].

6.4. Randomly assign the two task books to the students in the class

NOTE: To prevent contaminating factors related to where students are seated, distribute the task books homogeneously across the different parts of the class. For example, as you walk around the class give the PS-I task book to one student, then the direct instruction task book to the next student.

6.5. Once you have distributed the task books to all the students in the class, ask them to start working individually on the first activity.

6.5.1. Tell the students that they have 15 minutes for the first activity. Instructions are included in the paper sheets and no more general instructions are needed.

6.5.2. Tell them that you are available for any questions, but avoid giving students with any extra content other than what they have in the task books.

NOTE: Particularly for students solving the invention problem, avoid guiding them towards conventional solutions, because it can shortcut the development of their own knowledge[11]. Instead, we suggest three possible responses to student questions [11]: a) help them clarify their own processes by asking them to explain what they are doing; b) help them guide themselves with their intuition by asking them which country they think has more inequality than other countries; c) help them understand the goal of the activity by asking them to produce general indexes that would account for the differences they see, you can provide examples of other quantitative indexes (e.g., “the mean is an index to calculate the central value in a distribution”).

6.6. Once the 15 minutes for the first activity are over, ask students to advance to their corresponding second activity, for which they have to remove the clip or sticky note.

6.6.1. Tell them that they have 15 minutes for the second activity. Instructions are included in the paper sheets and no additional general instructions are needed. Tell them that you are available for any questions.

NOTE: Students have access to the content from the previous activity.

6.7. Once the 15 minutes are over, ask them to hand the completed material to you.

7. Administration of the lecture content

7.1. Reserve 40 minutes within one or several class periods to give the lecture about statistical variability to all students in the class.

NOTE: The protocol can be interrupted at any point during the lecture and can continue in the subsequent class session.

7.2. To give the lecture, follow the slides, which can be found at the following link: https://www.dropbox.com/sh/aa6p3hs8esyf5xa/AACTvpVlEbdEtLVfBIbe9j7aa?dl=0.

NOTE: The file includes animations to stagger the contents, comments with proposed explanations to give to students, and indications about the approximate time allocated for each explanation. The content and activities included are about the definition of variability, the use of different variability measures (range, interquartile range, variance, standard deviation, and coefficient of variation), the properties of those measures, and their advantages and disadvantages compared to each other and to other suboptimal solutions[13]. A further description of this proposed lecture can be found in **Appendix E**. The user can adapt these materials depending on different factors such as specific content to cover in class, preferred instruction principles, or different cultural expressions.

8. Completion of the curiosity questionnaire

8.1. At the end of the lecture, give students the Curiosity Scale from the Epistemic Related Emotions Questionnaire[40] (**Appendix F**) and give them 2 minutes to complete it. Remind students to write their identification number on the questionnaire before handing it back.

NOTE: In the literature, curiosity is often measured right after the invention activity and the corresponding control activities[14,17]. The protocol is flexible to this and other possible adaptations in this regard. For simplicity, we only included the measurement of curiosity at the end of the lesson because it is relevant to examining the longer-term effects of PS-I on curiosity, and because increased curiosity right after the invention activity can be partially explained by the fact that during the invention activity students receive less information than during alternative activities used as controls.

9. Administration of the learning post-test

9.1. In accordance with the teacher in each class, take 30 minutes in a class period to administer the post-test.

9.2. Distribute the post-test in **Appendix G** to the students. Ask them to work on it individually.

9.2.1. Give students 25 minutes to do the post-test. Instructions are included in the post-test and no additional general instructions are needed.

9.3. Once the 25 minutes are up, ask them to hand the post-test back to you.

10. Providing students with feedback and all learning materials

10.1. Make the materials used for this lesson available to students. The power-point slides, the materials for the two learning conditions, and the solutions for the pre-test and post-test are available in **Appendix H**.

11. Coding the data

11.1. Calculate the scores for the different scales in the questionnaires by adding together all the item scores within each questionnaire scale (see **Appendix B** for a summary of the questionnaire items in the proposed questionnaires).

11.2. Calculate the score for divergent thinking fluency by counting up all the appropriate responses given by each student in both items in the Alternative Uses Task[37].

NOTE: Other measures often coded from the Alternative Uses Task, such as flexibility, originality, and elaboration, might also be considered[36,37].

11.3. Calculate the score of the previous knowledge pre-test by first grading each item using the answer key in **Appendix I-1** and then adding together the scores for all of the items.

11.4. Calculate the different learning measures by first grading each item in the post-test using the answer key in **Appendix I-2** then adding together the scores for each learning measure: scores in items 1 to 3 for the procedural learning measure, scores in items 4-8 for the conceptual learning measure, and scores in items 9-11 for the transfer of learning measure.

NOTE: Other measures about the learning process such as the number of solutions produced by students during the invention problem or the correctness of the solutions in all problem-solving activities might be considered, but they will not be explained in this protocol.

12. Analysis of the data

Please note that references in this section refer to practical manuals on how to perform the analyses with SPSS and PROCESS software but other programs may also be used.

12.1. To evaluate the general efficacy of PS-I, compare the curiosity and learning scores of the PS-I condition versus the curiosity and learning scores of the control condition.

NOTE: As long as assumptions are fulfilled, we primarily recommend ANCOVA to control for predisposition of covariates. As a second option we recommend t-tests for independent groups and as a third option we recommend Mann-Whitney U tests[41]. No minimum sample size is required for these analyses, but considering the effect sizes in previous literature (d = .43)[21], a minimum sample of 118 students per group would be recommended to facilitate the identification of the effects as significant (two-tailed power analyses for differences between independent means, α = .05, β = .95,). Samples larger than 30 students per group would make it easier to meet the assumptions of normality for ANCOVA or t-tests[41].

12.2. To intuitively explore mediation effects (e.g., the mediation of curiosity on learning) and/or the moderating influence of predispositions, perform correlational analyses between the two learning conditions.

NOTE: As long as assumptions are fulfilled, we primarily recommend the use of Pearson correlations and as a second option we recommend Spearman correlations[42]. No minimum sample size is required for these analyses, but large samples (e.g., more than 30 students per group) would make it easier to fulfil the assumptions of normality needed for Pearson correlations. Possible moderation effects would be indicated by predisposition variables that have different correlation values in one learning condition versus the other. A possible mediation effect (e.g., the mediation of curiosity on learning) would be indicated if the mediating variable is correlated with the learning outcomes in at least one condition, and if the levels of this variable are different in one learning condition compared to the other (see results in Step 12.1).

12.3. To continue evaluating a mediation effect on learning and/or the moderating influence of students' predispositions, perform either mediation analysis, moderation analysis, or conditional process analysis (which combines mediation and moderation analysis) depending on the conceptual model to test[43], which would vary depending on the hypotheses chosen and/or the preliminary analysis in Step 12.2.

NOTE: Since these analyses are based on multiple regressions, and are therefore based on a fixed effect statistical approach, in order to make the results as generalizable as possible, we recommend a minimum sample size of 15 students per mediation variable included in the conceptual model, plus 30 students per moderation variable included in the model. Some programs such as PROCESS only allow the inclusion of a maximum of two moderating variables at one time. To incorporate more moderating variables, several analyses would need to be run changing the moderators included.

REPRESENTATIVE RESULTS:

This protocol was satisfactorily implemented in a previous study[23], with the exception of the measures of students' predispositions in terms of their sense of competence, mastery approach goals, metacognition, and divergent thinking.

To address these predispositions, this protocol includes measures that have been previously validated and that have shown high levels of reliability (**Table 1**).

Typical solutions generated by students in the invention problem of the PS-I condition can be seen in **Figure 3A-D**. Students do not usually produce the canonical solution of standard deviation. However, the sub-optimal solutions they do produce reveal reflection about relevant aspects of standard deviation (e.g., range, summing deviations, or averaging deviations). Previous research has shown that the variety of solutions in the initial problem in PS-I was associated with higher learning, regardless of the correctness of the response[44]. Nonetheless, it is important to note that the absence of response in this problem is not an indicator of students not benefiting from it, since students can critically reflect about the problem without producing a visible result.

A typical solution produced by students in the practice problem used in the control condition (**Figure 2**) is shown in **Figure 3 E**. These solutions are more homogeneous and in line with the canonical concept of standard deviation because it is a problem that was presented after they had studied the concepts and procedures in the Worked Example (**Appendix D-2**).

Figure 5 reproduces an example for reporting the general differences between PS-I and direct instruction in the experimental evaluation. It is based on results of a previous study that followed this protocol[23] in which students in the PS-I condition did not differ in procedural knowledge, transfer of knowledge, curiosity, or previous knowledge, but did differ in conceptual knowledge. **Figure 6** shows an example for reporting the moderating effect of one of the proposed student predispositions, metacognitive abilities. In this hypothetical example, students with lower metacognitive abilities learned more from direct instruction than from PS-I, while those with higher metacognitive abilities benefited more from PS-I than from direct instruction.

Figure 1: Invention Problem in the PS-I Condition.
In this problem[23] students in the PS-I condition are asked to invent quantitative indexes to measure inequality across the four countries. It is formulated with the technique of Contrasting Cases[11]: the countries show consistencies and variations regarding the relevant features, and these variations are easy to calculate. For example, Pinpanpun and Toveo have the same mean (5), same number of cases (7), same range (10), but different distribution.

Figure 2: Practice Problem in the Direct Instruction Condition.
In this problem[23] students in the direct instruction condition are asked to apply the concepts and procedures learned in the Worked Example.

Figure 3: Common Solutions in the Invention Problem and in the Practice Problem.
Images A-D show common solutions in the Invention Problem, which can be used in the posterior direct instruction phase to scaffold contents: (A) The range – easy to calculate, but does not

account for differences across all inhabitants-; (B) Range based measure - considers more inhabitants than the range as it becomes amplified when maximums values are repeated, but does not consider all values-; (C) Average of deviations - it accounts for differences across all inhabitants, but it is confusing because negative deviations subtract from positive deviations-; (D) Average of absolute deviations -a conceptually complete solution similar to the canonical solution of the standard deviation-; (E) A typical solution to the practice problem of the control condition. Students in this condition have already studied the Worked Example, and therefore most of them are able to reproduce and interpret correctly the canonical solutions of the standard deviation.

Figure 4: Design of Experimental Evaluation.
After the completion of the questionnaires and tests to measure students' predispositions, students are randomly assigned to the activities of the two learning conditions (all students remain in the same class). Once students complete these activities, all of them receive the same lecture about statistical variability. Curiosity and learning are measured at the end of the learning process.

Figure 5: Results about Efficacy of PS-I versus Direct Instruction.
The graphics display a typical result of the comparison between the PS-I condition and the direct instruction condition within each dependent variable, using data of a previous study that used this protocol[23]. The two bars in each graphic represent the means for the two conditions, while their corresponding error bars represent +/- 1 standard errors of those means. * indicates significant results at the .05 significance level.

Figure 6: Hypothetical Results about the Moderating Effects of Students' Predispositions
The graphics display an hypothetical result about the moderating effect of metacognitive abilities on the relative efficacy of PS-I to promote learning, in which PS-I is more effective than direct instruction only for students who report medium and high metacognitive abilities. Following recommendations in[43], the 16^{th} , 50^{th}, and 86^{th} percentiles have been used to respectively represent students with low, medium, and high metacogntive abilities.

Table 1: Proposed Constructs and Measures to Evaluate Students' Predispositions.
Five constructs about students' predispositions are proposed to be evaluated as moderators in the efficacy of PS-I. A proposed measure for each construct is described regarding the number of items, description of the items, and evidence about validity and reliability.

Table 2: Proposed Constructs and Measures to Evaluate the efficacy of PS-I.
The proposed instruments to measure curiosity and three types of learning (procedural, conceptual, and transfer) are described, including information about number of items, description of the items, and evidence about validity and reliability.

DISCUSSION:
The aim of this protocol is to guide researchers and educators in the implementation and evaluation of the PS-I approach in real classroom contexts. According to some previous

experiences, PS-I can help promote deep learning and motivation in students [19,21,24], but there is a need for more research about its efficacy in students with different abilities and motivational predispositions[14,27]. More specifically, using this document, educators can follow a PS-I implementation protocol for a statistics class designed according to the most widely-accepted principles in the PS-I literature[11,13,20,50] (Steps 6-7). Additionally, educators and researchers can follow an embedded experimental evaluation about the efficacy of this implementation in students with different motivational and/or cognitive predispositions (all Steps). This experimentation does not conflict with the educational principles of equality of opportunities, free consent to participate, or respecting student confidentiality, nor is it necessary to use any new technologies.

The protocol is flexible and may be modified or applied according to new research or educational needs. Nevertheless, as described in this document, the protocol allows the evaluation of the efficacy of PS-I in terms of curiosity and different types of learning, including learning measures that require deep learning, such as conceptual knowledge and transfer of knowledge, as well as learning measures that do not necessarily require deep learning, such as procedural knowledge. Both motivation and deep learning are significant concerns for all instructors. STEM course designers are especially concerned with these topics as a large proportion of students have difficulties understanding those courses[1-3] and experience various motivational issues[4,5]. The protocol also provides guidance for the evaluation of the efficacy of PS-I in students in terms of some cognitive and/or motivational predispositions, which are also a concern in STEM education, and in the relative efficacy of PS-I. The predispositions proposed in the protocol include previous academic knowledge, mastery-approach goals, sense of competence learning the subject, metacognition, and divergent thinking.

Examples of modification to the protocol based on ideas proposed in the literature include increasing the number of problems in the conditions[15], giving students more time for problem exploration[44], and including different variables to account for mediational learning processes[14,15,24]. The protocol is also flexible about the application of the different steps over different class sessions. Each step can be performed in the same class period as the previous step, and researchers and educators can decide how to organize the steps to their own convenience.

Nevertheless, a critical factor for the evaluation is that students collaborate in respecting the evaluation rules. For example, in some steps they are supposed to work individually so that possible interactions between them do not contaminate the results. In order to achieve that, it is important for students to be informed about the procedures, and for them to be equally involved in the learning activities regardless of whether they want to participate in the experimental evaluation or not [32], as described in Step 1 of the protocol. For the activities that require individual work, we also recommended ensuring that there are spaces left between students.

In summary, this protocol may be useful in making PS-I and its experimental evaluation more accessible to educators and researchers, providing them with materials and guidance, giving them the flexibility to apply it according to their research and educational needs, and proposing

analysis options that adapt to different sample sizes. However, one possible limitation here might be the time required to complete the questionnaires and tests about student predispositions. When the user is interested in evaluating these predispositions but there is no available time to do so during class, these questionnaires could be completed as an assignment outside class. A second limitation is the potential measurement error of some of the proposed predisposition measures that are not specifically contextualized in the learning of variability measures, but rather in general learning (metacognition and divergent thinking) or general statistics learning (mastery approach goals and sense of competence). This error should be considered as a potential limitation of any studies conducted with this protocol. A final limitation is that the previous knowledge pre-test and the learning post-test are not validated measures in the previous literature so far since the content of the implementation is very specific and validated measures for them are not available. However, it is expected that the future implementation of this protocol will advance their validation.

On similar lines, future application of the protocol will also define new research needs and new variations to be applied. Having the protocol as a common source may contribute to provide a certain systematic structureacross different studies. In addition, as long as the educators find the experimental evaluation of this protocol compatible with their educational practice, this protocol may encourage involvement of educators with PS-I research, which would mean a broader professional perspective in the research process and better access to samples [32].

ACKNOWLEDGMENTS:

This work was supported by a project of the Principality of Asturias (FC-GRUPIN-IDI/2018/000199) and a predoctoral grant from the Ministry of Education, Culture, and Sports of Spain (FPU16/05802). We would like to thank Stephanie Jun for her help editing the English in the learning materials.

DISCLOSURES:

The authors have nothing to disclose.

REFERENCES:

1 Silver, E. A., Kenney, P. A. *Results from the seventh mathematics assessment of the National Assessment of Educational Progress*. (National Council of Teachers of Mathematics, 2000).

2 OECD. *PISA 2015 Results (Volume I): Excellence and Equity in Education*. PISA, OECD publishing. Paris, France (2016).

3 Mallart Solaz, A. La resolución de problemas en la prueba de Matemáticas de acceso a la universidad: procesos y errores. (2014).

4 García, T., Rodríguez, C., Betts, L., Areces, D., González-Castro, P. How affective-motivational variables and approaches to learning predict mathematics achievement in upper elementary levels. *Learning and Individual Differences.* **49** 25-31 (2016).

5 Lai, Y., Zhu, X., Chen, Y., Li, Y. Effects of mathematics anxiety and mathematical metacognition on word problem solving in children with and without mathematical learning difficulties. *PloS one.* **10** (6), e0130570 (2015).

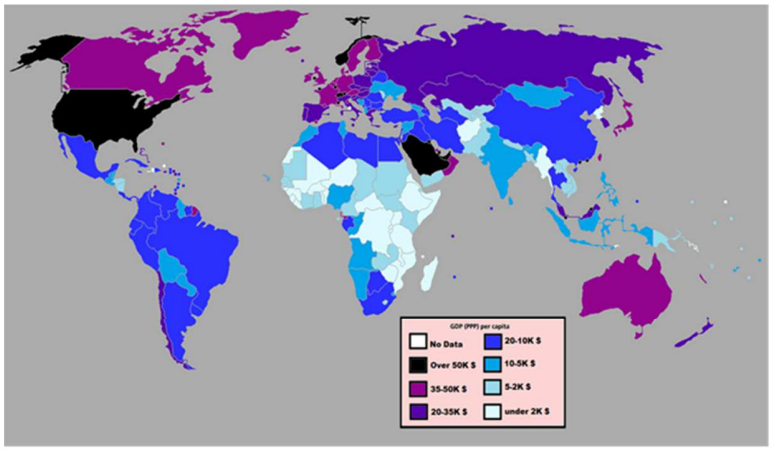
6 Ma, X., Xu, J. The causal ordering of mathematics anxiety and mathematics achievement: a longitudinal panel analysis. *Journal of Adolescence.* **27** (2), 165-179 (2004).
7 Kapur, M. Productive Failure in Learning Math. *Cognitive science.* **38** (5), 1008-1022 (2014).
8 Kirschner, P. A., Sweller, J., Clark, R. E. Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist.* **41** (2), 75-86 (2006).
9 Stockard, J., Wood, T. W., Coughlin, C., Khoury, C. R. The Effectiveness of Direct Instruction Curricula: A Meta-Analysis of a Half Century of Research. *Review of educational research.* **88** (4), 479-507 (2018).
10 Clark, R., Kirschner, P. A., Sweller, J. Putting students on the path to learning: The case for fully guided instruction. *American Educator* (2012).
11 Schwartz, D. L., Martin, T. Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and instruction.* **22** (2), 129-184 (2004).
12 Loibl, K., Rummel, N. The impact of guidance during problem-solving prior to instruction on students' inventions and learning outcomes. *Instructional Science.* **42** (3), 305-326 (2014).
13 Kapur, M., Bielaczyc, K. Designing for Productive Failure. *Journal of the Learning Sciences.* **21** (1), 45-83 (2012).
14 Glogger-Frey, I., Fleischer, C., Grueny, L., Kappich, J., Renkl, A. Inventing a solution and studying a worked solution prepare differently for learning from direct instruction. *Learning and Instruction.* **39** 72-87 (2015).
15 Glogger-Frey, I., Gaus, K., Renkl, A. Learning from direct instruction: Best prepared by several self-regulated or guided invention activities? *Learning and Instruction.* **51** 26-35 (2017).
16 Likourezos, V., Kalyuga, S. Instruction-first and problem-solving-first approaches: alternative pathways to learning complex tasks. *Instructional Science.* **45** (2), 195-219 (2017).
17 Lamnina, M., Chase, C. C. Developing a thirst for knowledge: How uncertainty in the classroom influences curiosity, affect, learning, and transfer. *Contemporary educational psychology.* **59** 101785 (2019).
18 Loibl, K., Rummel, N. Knowing what you don't know makes failure productive. *Learning and Instruction.* **34** 74-85 (2014).
19 Weaver, J. P., Chastain, R. J., DeCaro, D. A., DeCaro, M. S. Reverse the routine: Problem solving before instruction improves conceptual knowledge in undergraduate physics. *Contemporary educational psychology.* **52** 36-47 (2018).
20 Loibl, K., Roll, I., Rummel, N. Towards a Theory of When and How Problem Solving Followed by Instruction Supports Learning. *Educational psychology review.* **29** (4), 693-715 (2017).
21 Darabi, A., Arrington, T. L., Sayilir, E. Learning from failure: a meta-analysis of the empirical studies. *Etr&D-Educational Technology Research and Development.* **66** (5), 1101-1118 (2018).
22 Chen, O. H., Kalyuga, S. Exploring factors influencing the effectiveness of explicit

instruction first and problem-solving first approaches. *European Journal of Psychology of Education.* 10.1007/s10212-019-00445-5 18 (2019).
23 González-Cabañes, E., García, T., Rodríguez, C., Cuesta, M., Núñez, J. C. Learning and Emotional Outcomes after the Application of Invention Activities in a Sample of University Students. *Sustainability.* **12** (18), 7306 (2020).
24 Schwartz, D. L., Chase, C. C., Oppezzo, M. A., Chin, D. B. Practicing Versus Inventing With Contrasting Cases: The Effects of Telling First on Learning and Transfer. *Journal of educational psychology.* **103** (4), 759-775 (2011).
25 Chase, C. C., Klahr, D. Invention Versus Direct Instruction: For Some Content, It's a Tie. *Journal of Science Education and Technology.* **26** (6), 582-596 (2017).
26 Newman, P. M., DeCaro, M. S. Learning by exploring: How much guidance is optimal? *Learning and Instruction.* **62** 49-63 (2019).
27 Belenky, D. M., Nokes-Malach, T. J. Motivation and Transfer: The Role of Mastery-Approach Goals in Preparation for Future Learning. *Journal of the Learning Sciences.* **21** (3), 399-432 (2012).
28 Bergold, S., Steinmayr, R. The relation over time between achievement motivation and intelligence in young elementary school children: A latent cross-lagged analysis. *Contemporary educational psychology.* **46** 228-240 (2016).
29 Mazziotti, C., Rummel, N., Deiglmayr, A., Loibl, K. Probing boundary conditions of Productive Failure and analyzing the role of young students' collaboration. *NPJ science of learning.* **4** 2-2 (2019).
30 Stiglitz, J. E. Las limitaciones del PIB. *Investigacion y ciencia.* **October** (529), 26-33 (2020).
31 Holmes, N. G., Day, J., Park, A. H., Bonn, D., Roll, I. Making the failure more productive: scaffolding the invention process to improve inquiry behaviors and outcomes in invention activities. *Instructional Science.* **42** (4), 523-538 (2014).
32 Herreras, E. B. La docencia a través de la investigación-acción. *Revista Iberoamericana de Educación.* **35** (1), 1-9 (2004).
33 Schau, C., Stevens, J., Dauphinee, T. L., Delvecchio, A. The development and validation of the survey of attitudes toward statistics. *Educational and Psychological Measurement.* **55** (5), 868-875 (1995).
34 Elliot, A. J., Murayama, K. On the measurement of achievement goals: Critique, illustration, and application. *Journal of educational psychology.* **100** (3), 613-628 (2008).
35 Schraw, G., Dennison, R. S. Assessing metacogntive awareness. *Contemporary educational psychology.* **19** (4), 460-475 (1994).
36 Guilford, J. P. The nature of human intelligence. (1967).
37 Zmigrod, L., Rentfrow, P. J., Zmigrod, S., Robbins, T. W. Cognitive flexibility and religious disbelief. *Psychological Research-Psychologische Forschung.* **83** (8), 1749-1759 (2019).
38 Wilson, S. Divergent thinking in the grasslands: thinking about object function in the context of a grassland survival scenario elicits more alternate uses than control scenarios. *Journal of Cognitive Psychology.* **28** (5), 618-630 (2016).
39 Autin, F., Croizet, J.-C. Improving working memory efficiency by reframing metacognitive interpretation of task difficulty. *Journal of experimental psychology: General.* **141** (4), 610 (2012).
40 Pekrun, R., Vogl, E., Muis, K. R., Sinatra, G. M. Measuring emotions during epistemic

activities: the Epistemically-Related Emotion Scales. *Cognition and Emotion.* **31** (6), 1268-1276 (2017).
41 Pallant, J. Statistical techniques to compare groups. In *SPSS survival manual*. 211-329, McGraw-Hill Education, UK (2013).
42 Pallant, J. Statistical techniques to explore relationships among variables. In *SPSS survival manual*. 125-149, McGraw-Hill Education, UK (2013).
43 Hayes, A. F. *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford publications. New York, NY (2017).
44 Kapur, M. Productive failure in learning the concept of variance. *Instructional Science.* **40** (4), 651-672 (2012).
45 Nolan, M. M., Beran, T., Hecker, K. G. Surveys Assessing Students' Attitudes Toward Statistics: A Systematic Review of Validity and Reliability. *Statistics Education Research Journal.* **11** (2) (2012).
46 Schraw, G., Dennison, R. S. Assessing metacognitive awareness. *Contemporary educational psychology.* **19** (4), 460-475 (1994).
47 Dumas, D., Dunbar, K. N. Understanding Fluency and Originality: A latent variable perspective. *Thinking Skills and Creativity.* **14** 56-67 (2014).
48 Roberts, R. *et al.* An fMRI investigation of the relationship between future imagination and cognitive flexibility. *Neuropsychologia.* **95** 156-172 (2017).
49 Chamorro-Premuzic, T. Creativity versus conscientiousness: Which is a better predictor of student performance? *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition.* **20** (4), 521-531 (2006).
50 Kapur, M. Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educational Psychologist.* **51** (2), 289-299 (2016).

Figure 1. Invention Problem in the PS-I Condition

Click here to access/download;Figure;Fig 1.pdf

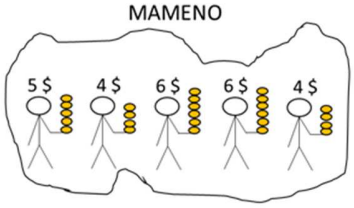


This image shows the average income (GDP per capita) of the countries in 2012. We can see that, for example, Spain and Russia have a similar average income. However, only with the average income, can we have an idea of the wealth of the inhabitants?

Below you can see 4 imaginary countries. Although the average income is the same in all of them, it is easy to see that they differ on how inequality strongly affects the wealth of their inhabitants.
How can we measure inequality?

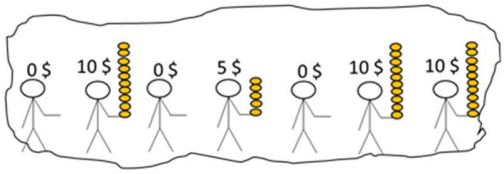
Try to design a mathematical index that would help us to measure and compare inequality in these 4 countries. You can give multiple answers. There is no one unique valid solution!

MAMENO



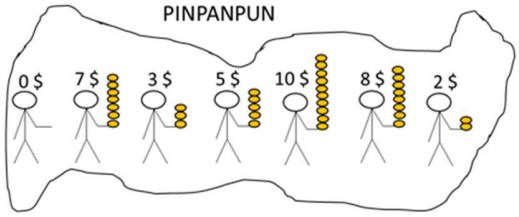
Average Income: (5+ 4 + 6 + 6 + 5) / 5 = **5**

TOVEO



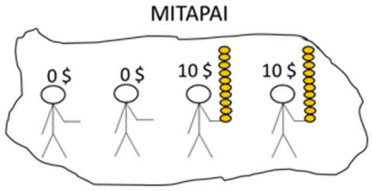
Average Income: (0 + 10 + 0 + 5 + 0 +10 + 10) / 7 = **5**

PINPANPUN



Average Income: (0 + 7 + 3 + 5 + 10 + 8 + 2) / 7 = **5**

MITAPAI



Average Income: (0 + 0 + 10 + 10) / 4 = **5**

Calculate the range, the interquartile range, and the standard deviation for the income distribution in Tristras. Discuss the affordances and limitations of how these three measures allow to observe higher or lower inequality in Tristras versus the four countries you have seen in the previous activity.

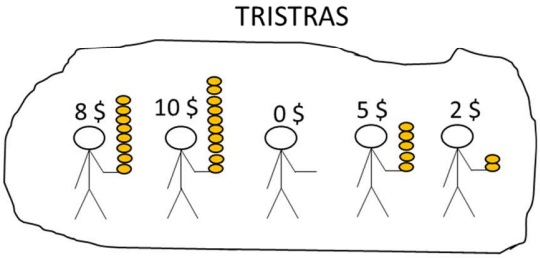
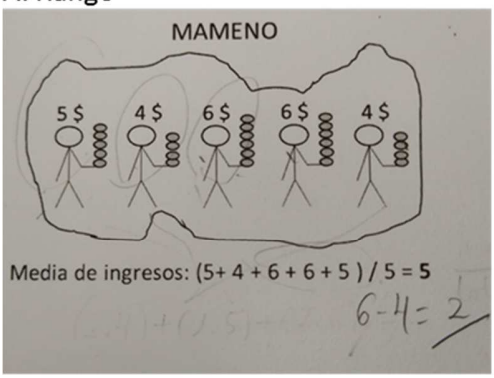


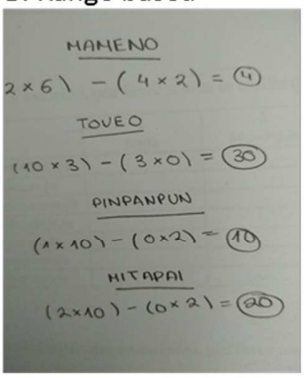
Figure 3: Common Solutions in the Invention Problem and in the Practice Problem.

Click here to access/download;Figure;Fig 3.pdf

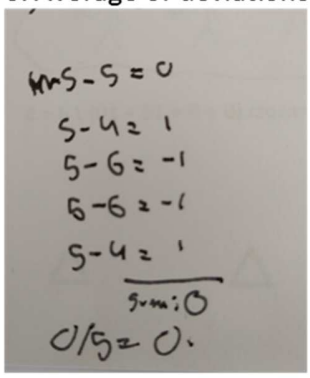
A. Range



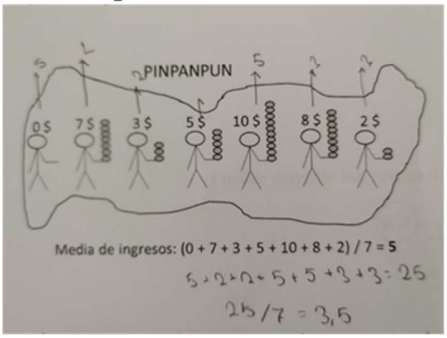
B. Range based



C. Average of deviations



D. Average of absolute deviations



E. Standard deviation

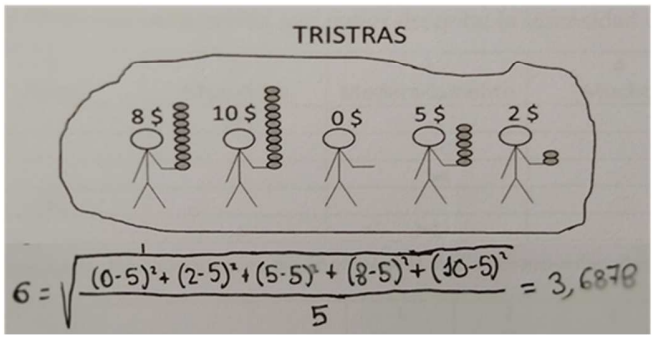


Figure 4: Design of Experimental Evaluation

Click here to access/download;Figure;Fig 4.pdf

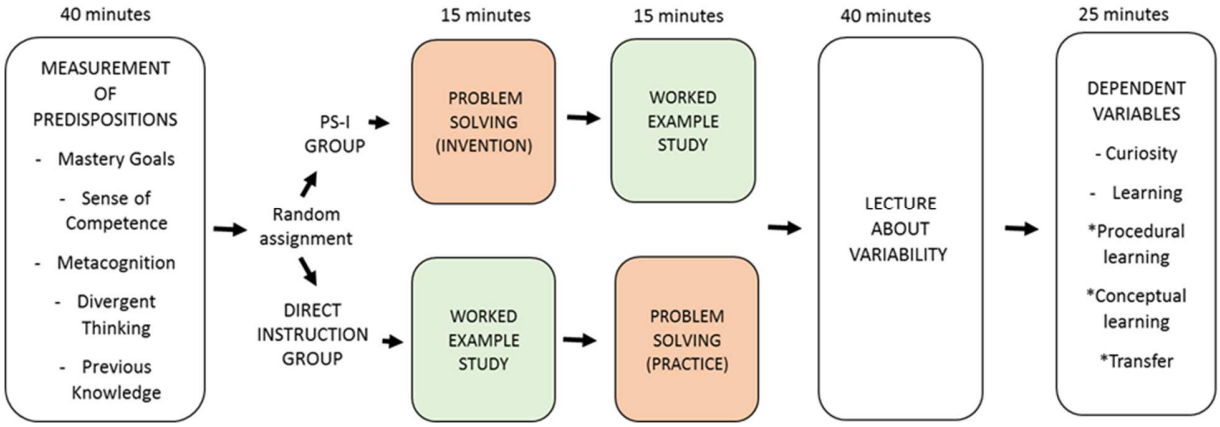


Figure 5: Results about Efficacy of PS-I versus Direct Instruction

Click here to access/download;Figure;Fig 5.pdf

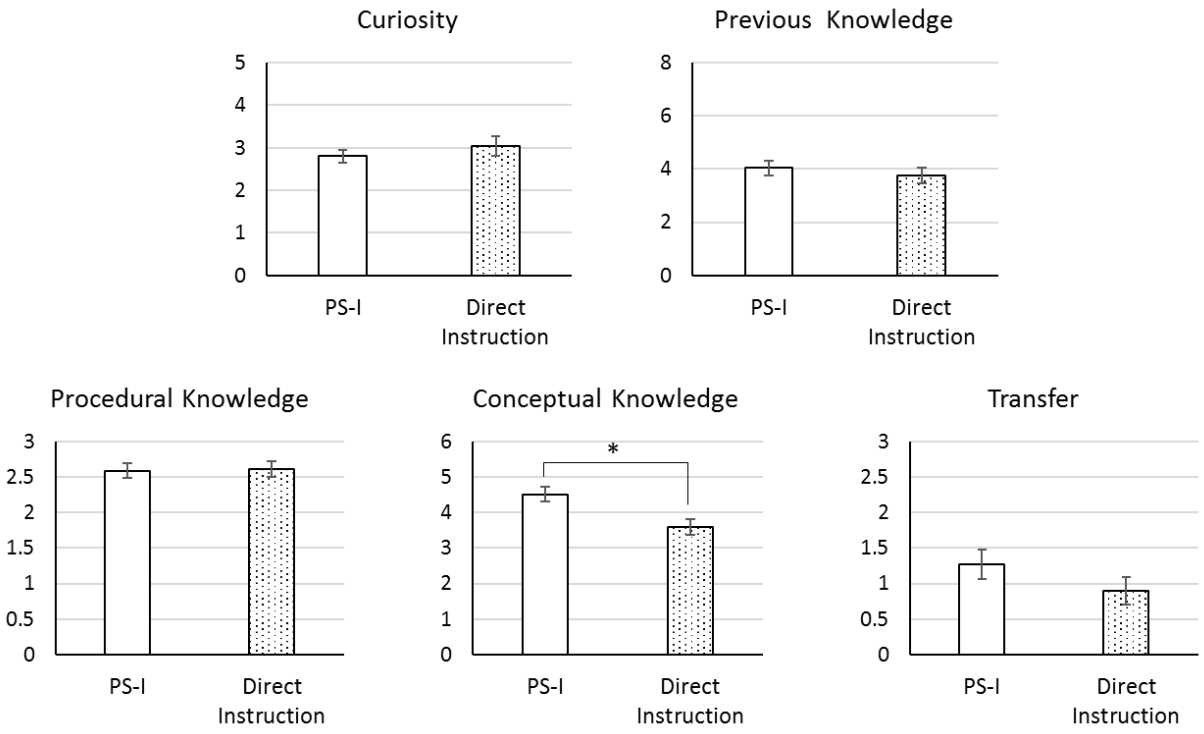


Figure 6: Figure 6: Hypothetical Results about the Moderating Effects of Students' Predispositions

Click here to access/download;Figure;Fig 6.pdf

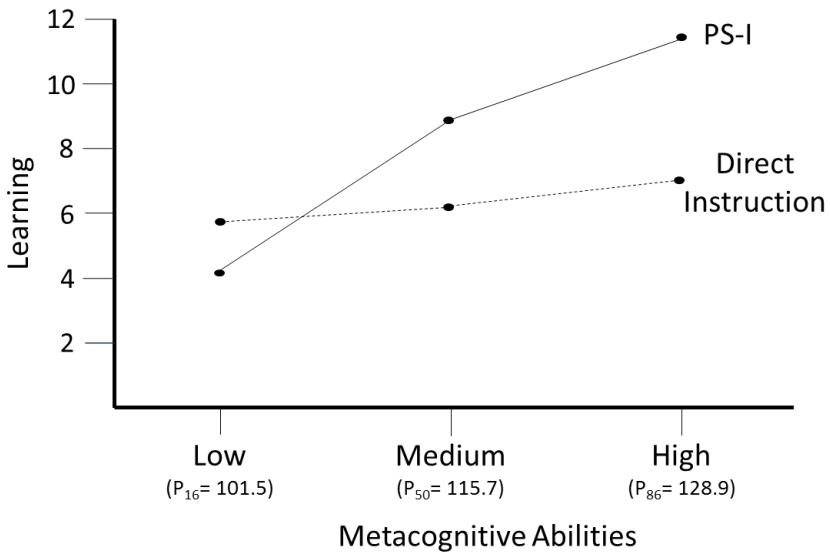


Table 1: Proposed Constructs and Measures to Evaluate Students' Predispositions.

Click here to access/download;Table;Table 1.xls

Construct
Sense of Competence
Mastery Approach Goals
Metacognitive Regulation
Divergent Thinking
Previous Academic Knowledge

Measure and Description
The Cognitive Competence Scale in the Survey of Attitudes towards Statistics (SATS-28)[33] can be used (Appendix B2). It is composed of 6
The Mastery Approach Scale in the Achievement Goal Questionnaire-Revised[34] can be used (Appendix B3). It is composed of 3 items that ask students how much they agree with statements about having learning goals that focus on personal learning (e.g., "I am striving to understand the content of this course as thoroughly as possible"). It has shown internal, convergent and predictive validity, and high internal reliability (α = .84)[34].
The Regulation of Cognition Scale of the Metacognitive Awareness Inventory[46] can be used (Appendix B4). It consists of 35 items that ask students how typical it is for them to use different metacognitive strategies (e.g., "I reevaluate my assumptions when I get confused"). It has shown internal and predictive validity, and high reliability (α = .88)[46].
The Fluency score from the Alternative Uses Task[36] can be used. It consists of presenting students with several objects (e.g., a paper clip), and asking them to provide as many uncommon uses for each object within a given time. It is a reliable score (H = .631) that has been internally validated [47] and has shown predictive validity in versions with different extensions, varying between 1 to 20 objects presented, and between 1 to 3 minutes given for each object[37,48,49]. For time restrictions within educational settings, a short version of two objects and two minutes per object[37] is proposed in this protocol.
To adapt to the specific contents covered in this protocol, a learning pre-test has been adapted (Appendix C) from a reliable (α =.75) pre-test used in a previous study [7]. It consists of 5 items that ask students about central tendency measures that are relevant to the assimilation

Table 2: Proposed Constructs and Measures to Evaluate the efficacy of PS-I

Click here to access/download;Table;Table 2.xls

Construct
Curiosity
Learning (procedural, conceptual, and transfer)

Measure and Description
The Curiosity Scale in the Epistemically-Related Emotions Questionnaire[40] can be used (Appendix F). It consists of three items that ask students to rate the intensity they felt curious, interested, and inquisitive. It has shown internal and predictive validity, and high reliability (α = .88)[40]
To evaluate learning about the specific variability contents covered in this protocol, a learning post-test has been adapted (Appendix G) from a reliable (α =.84) post-test used in a previous study[7]. It consists of 12 items: three items referred to procedural learning (e.g., item 1 where students have to calculate the standard deviation), six items referred to conceptual learning (e.g., item 4 where students have to reason about components of the standard deviation formula), and three items referred to transfer (e.g., item 10 where students have to

Click here to access/download
Table of Materials
JoVE_Materials.xls

Rebuttal Letter

Click here to access/download;Rebuttal Letter;RESPONSE TO REVIEWERS.docx

Manuscript ID: JoVE62138R1
Title: Problem-Solving Before Instruction (PS-I): A protocol for assessment and intervention in students with different abilities

Dear editors,

We acknowledge the comments and feedback from you and the reviewers, which have helped improve the quality of the video and manuscript.

As general clarifications, the introduction of the manuscript has been reformulated to address the different suggestions of the reviewers while having in mind the word limit. We have tried to make it clearer that the present manuscript describes a protocol, rather than an empirical study with results. Also, we have specified that the protocol is based on a previous study and included representative results from this study.

Below, we provide a detailed response to the comments and suggestions made by each reviewer of the manuscript. At the end we provide responses to the review of the video.

RESPONSES TO REVIEWER 1

Manuscript Summary:
The video and accompanying manuscript, "Problem-Solving Before Instruction (PS-I): A protocol for assessment and intervention in students with different abilities," present a clear protocol for conducting a lesson on statistical variability, using the PS-I strategy, at the undergraduate level. In this strategy, students explore a problem on the target concept before receiving direct instruction on the concept. Students in the PS-I condition are asked to invent solutions using contrasting cases to a standard deviation problem before studying worked examples of problem solutions and hearing a lecture. For researchers wishing to re-establish a basis for this procedure's efficacy, the research steps and materials are included. In this iteration, a comparison group in the same classroom studies the worked example problem solutions before practicing solving an isomorphic problem and then hearing the lecture. The steps are described with clarity, instructions are provided, and timing is suggested. All materials for implementation and assessment are given in the Appendix. Data analyses are recommended. A major asset of this work is in the authors' consideration of the important issue of students' sense of competence, motivational orientation, and metacognitive awareness and these predispositions' relationship to successful problem-solving before or after instruction. Another measurement of individual differences in response to the strategy, divergent thinking, is included as a possible moderator, as is prior knowledge of central tendency. Three learning outcomes are assessed on the post-test: procedural knowledge, conceptual knowledge, and transfer of knowledge. In addition, curiosity is measured as a mediator of the PS-I approach. Modifications are offered to educators wishing only to proceed with the implementation and not experimental evaluation. In sum, this is a valuable contribution that will facilitate educators' use of the PS-I strategy and establish a consistent basis for further research on its usage in STEM.

Many thanks for your comments.

Minor Concerns:
*** In many recent PS-I studies (e.g., Nachtigall, Serova, & Rummel, 2020; Weaver et al., 2018) effort has been made to keep the problem-solving and instructional materials equivalent across conditions. Judging from the task books in the present proposed design, the Direct Instruction condition sees a practice problem that the PS-I group never sees. I understand that this problem is isomorphic to the income inequality problem of the PS-I condition; however, I wonder if it offers an enhanced near-transfer opportunity for the DI students of which the PS-I students do not have the benefit. This isomorphic problem offers the DI students the opportunity to recapitulate the procedures, affordances, and limitations just described in the worked example. Why not direct the PS-I students to try to do the same? If concerns about time-on-task are at issue (since the protocol is - very practically - designed to take place within one class) it should be explained that while the worked example is the same between conditions, the practice problem is only seen by the Direct Instruction group. Figure 6 in the manuscript makes it appear (due to the color-coding, perhaps) that the problem solving phases of the experiment are equivalent. Whereas the worked example study is the same, the problem solving phase is not, strictly speaking.**

As suggested, we have specified that the practice problem is only presented to the direct instruction group. The aim of doing so is to compensate for the extra time invested by the PS-I group in the invention problem (lines: 225-226-219, and 471-473). We have also included a clarification about the difficulty of PS-I designs to control for equivalence in terms of both time and materials (lines: 473-480). Finally, we have also highlighted the option to adapt the protocol according to research needs which includes designs that are more equivalent in terms of materials presented (lines: 227-234).

*** The Discussion section of the manuscript mentions student collaboration but there is nowhere in the protocol or video where collaboration is suggested. Rather, it seems designed to be implemented with students working individually. This should be clarified in the manuscript.**

This was an error in our English expression. We have clarified that we originally used “collaboration” in the sense of motivation and involvement (that the general group of students respect the rules of the protocol evaluation, for example, not speaking with each other in sections where it is crucial that they work individually, such as the posttest) (lines 826-828).

*** Section 2 of the manuscript protocol could also mention the possibility of using the feature of learning management systems that randomly assigns students to groups (i.e., experimental conditions). That way the students would not have to keep track of their own identification number and a research assistant could apply the number to the students' work after they turned it in. This would also present the opportunity to keep students blind to knowing that they were being assigned to different learning conditions. (as long as time was equal across conditions and packets were individualized). It would also allow students to turn in their work and questionnaires with their names on them so that demographic data could later be attached to their performance and survey data by a research assistant or institutional review officer.**

A note has been added to include this suggestion (lines 367-369).

*** In section 7 of the protocol, it should be specified whether or not instructors are encouraged to modify the PowerPoint lecture. If not, there are some nonstandard usages and Spanish punctuation therein that may need to be adapted for a native English-speaking audience.**

It has been specified that instructors can modify the power-point lecture (lines 562-563).

*** The use of the word "part" in two instances in the manuscript does not easily translate into English. I had trouble understanding whether this meant students who "possess" or "do not possess" predispositions (line 69) or "come from" or "do not come from" lower academic levels (line 144) until its use was contextualized at the end of the paragraph by a concrete example (lines 145-147).**

It has been changed using to the suggested words (lines 75, 245).

*** Many lines of the manuscript are highlighted without explanation.**

The highlighting has been eliminated.

RESPONSES TO REVIEWER 2

This was a nicely written and organized overview of a methodology to implement productive failure/invention activities in a college classroom (but may also be used at other age levels). The level of detail seemed perfect, and the procedures were clear and justified by prior research. I have two general comments, and several more specific comments.

Many thanks for your comments.

GENERAL COMMENTS

A key issue with this protocol is the conceptualization of the direct instruction (control) condition. It is great that the activities are exactly the same between control and PS-I conditions, and just the order of activities is changed. This allows much better experimental comparison and control relative to some prior studies. However, it is debatable as to whether the control condition is a true direct instruction condition. For example, why do students in the control condition NOT receive the lecture on standard deviation before the worked example and activity? I realize this may be a logistic issue, in that it is hard to give a lecture to one group of students and not another. But it is also not a true direct instruction condition - this is not how a teacher typically would teach (i.e., with a worked example in a packet, then practice, and THEN a lecture). In addition, a similar methodology (worked example before instruction) has been treated in the literature as a type of scaffolded productive failure/invention activity (e.g., Glogger-Frey et al., 2015; Newman & DeCaro, 2019). Although the framing is certainly debatable, these studies showed better performance for students in the worked example-first condition (although there were some moderators to these effects, such as the presence of a pretest). At the very least, the similarities/differences between the proposed procedure (worked examples, practice, lecture) and prior studies (e.g., lecture then practice, or worked examples then lecture) need to be discussed and justified. In addition, I would suggest to those reading this protocol that, if they are able, a lecture-first method is also an option and advisable.

The option to use the lecture as a first step in the control condition has been suggested (lines 227-229). We have also clarified what we refer by "direct instruction" as an approach that can include starting with worked examples (lines 78-82). Finally, similarities and differences between the proposed control condition and other potential adaptations (lecture first, or worked example followed by lecture) have been discussed, providing references from previous literature (lines 229-234).

Second, the instructions to reviewers states that the efficacy of the protocol must be demonstrated. There was not a lot of detail given about the data collected using this protocol (e.g., sample, significance of results). There was also no actual data given on the survey measures that are a big part of this protocol. If this journal is looking for evidence of efficacy with regards to actual data, then this

aspect is somewhat missing from this paper. If the journal is looking for efficacy in terms of ability to implement this protocol, then the paper is great.

To provide more references of real data after the application of the protocol, we have changed the representative results in Figure 5 to display real results, specifically those by González-Cabañes, García, Rodríguez, Cuesta, & Núñez (2020).

Lines 86-88: It should be noted here that this procedure (i.e., instruction that builds upon student solutions) is not always done (e.g., in the invention literature).

This idea is now described in more detail in lines 175-179.

Lines 102-104: It should be noted that the findings for interest and engagement are mixed in the literature (e.g., Weaver et al., 2018).

This aspect has been noted in lines 117-120.

Lines 116-118: Similar to my first general comment above, I don't think that Glogger-Frey et al. (2015) would consider the worked examples first condition as a direct instruction method. Be careful about how you are conceptualizing these findings.

It has been clarified what we refer as direct instruction (lines 78-82).

Line 240: Assignment of participants to a random number could also be done after the fact (after all data has been combined). It is certainly much easier to combine classroom data by name than expect students to remember their number over several sessions.

It has been clarified that the procedure we used was aimed to maintain the anonymity of students´ responses (lines 361-369).

Line 410: Why not give the curiosity scale after the activity? Theory would suggest that curiosity would be higher after invention, not after the lecture has been completed. Prior research is consistent with this too.

This option has been mentioned as a potential adaptation, while the inclusion of the measurement of curiosity at the end of the lesson in the present study has been justified (lines 572-578).

Line 438: Is this the standard method for scoring the Alternative Uses task? I thought there were standardized procedures that people typically follow, such as looking for commonly given answers and not giving credit for those. I could be wrong.

This would correspond to other measure taken from the Alternative Uses Task, specifically the measure of originality. For simplicity we only included the measure that corresponds to divergent thinking, fluency, but we have included a note specifying that other measures from this test might be used (lines

608-609).

Line 465: You mean 30 students per condition, right? If you divide 30 students into two conditions, you would only have 15 in each, which is not nearly enough. Also, what is the recommended sample size based on prior effect sizes in the literature? You could at least mention what the effect size is that could be used in a power analysis.

We have specified 30 students per condition, and we have also provided recommendations of sample sizes according to power analysis (lines 638-643 598-602).

Line 498: I disagree with using the total score for mediation/moderation analyses. The inclusion of the procedural knowledge subscale will only dilute the findings.

In coherence with this comment, we have eliminated this recommendation about using the total score for learning (lines 677-679).

Line 531: If you have conducted this study already, do you not have actual data you could report? I don't know if that's necessary for this journal, it just seemed odd to report some actual data and some hypothetical data.

This paragraph referred to hypothesis has been deleted (lines 713-715). Also, the referred figure has been deleted, following recommendations of a posterior comment.Finally, we have changed the representative results in Figure 5 to display real results by González-Cabañes, García, Rodríguez, Cuesta, & Núñez (2020).

Figure 4: This figure is very confusing, and I don't think it is needed. I also don't know that prior studies would justify all of these links. At the least, more research is needed. I know that it is hypothetical, but I also don't know that it's necessary to include here.

This Figure has been eliminated. Please note that the order of the rest of the Figures have changed because of this and other figure deletions commented later (e.g. Figure 4 now corresponds to previous Figure 6).

Figure 5: I also don't know if Figure 5 is helpful or necessary, or justified. Are these factors really all mediated by curiosity? Would there not be direct effects on learning (e.g., for competence)? I don't know that there is enough prior literature on how all these constructs relate to support this model. I realize this is hypothetical, but I'm not sure why it's needed, either.

The Figure has been eliminated. Please note that the order of the rest of the Figures have changed because of this and other figure deletions (e.g. Figure 5 now corresponds to previous Figure 7).

Figure 7: Is this the actual from the study reported earlier? If so, which effects are significant? Is general learning just the other three subscales summed?

We have changed this Figure (Now Figure 5) to display real results by González-Cabañes, García, Rodríguez, Cuesta, & Núñez (2020). Significant results were marked with * and a better description can be seen in the referred article. The “General learning” measure has been eliminated according to a previous comment.

Figure 8: The title of this figure should include the word "Hypothetical," as that can be easily missed. Given that it comes after Figure 7, which reports actual data (I think), it makes Figure 8 even more likely to be interpreted as including actual data.

As suggested, this have been clarified adding the word “Hypothetical” to this Figure (now Figure 6)**.**

RESPONSES TO REVIEWER 3

Manuscript Summary:
The manuscript "Problem-Solving Before Instruction (PS-I): A protocol for assessment and intervention in students with different abilities" suggested the use of PS-I intervention for students' cognitive and affective development. This manuscript includes the protocol and PS-I intervention example, which can increase students' conceptual understanding through the process of problem solving.

Major Concerns:
The protocol is well-designed. However, there are many issues to make me concern about publishing this manuscript in JoVe.

Many thanks for your comments.

1) There is no conceptual framework of Problem-Solving Before Instruction.

The introduction has been reformulated to improve the conceptual framework for PS-I (lines 87-104).

2) No participants or students backgrounds explained. For example, who are the targeted students? Problem solving should consider students' grade/year level at least, because of strong relationships between students' grade level/cognitive level and their problem solving results. What's the sample size?

Justification about the targeted students population has been added (lines 210-211). Also, it has been added a recommendation for the sample size according to power analysis (lines 637-641).

3) No analytic backgrounds. How did you come up with Figures 4 and 5? To show these relationships between items, analytic supports should be indicated. What statistical or theoretical analysis did you use?

These Figures have been eliminated according to these and other comments from other reviewers.

4) Specific statistical results should be indicated for Figure 7 (ex., M, SD, Cohen's d, 95% CI, etc.) I cannot see whether the impact of the intervention was statistically significant or not.

This Figure (now Figure 5) has been changed to display real results by González-Cabañes, García, Rodríguez, Cuesta, & Núñez (2020). This reference is mentioned to give access to the statistical data. Additionally, significant differences have been marked with "*".

5) You also need to provide supportive proof of how the intervention problem was developed. Why

did you pick the GDP or income related contents? Why is that meaningful to students' learning? How does it mean to math education?

It has been justified the use of the GDP to contextualize the invention problem (lines 211-213, and 456-460).

6) Statistical results of Pre-and Post-tests are required.

Figure 5 now includes data about both pre-test and post-test from the study by González-Cabañes, García, Rodríguez, Cuesta, and Núñez (2020)

7) The instruments you provided had great Cronbach's alpha to use. However, I am questioning if the questionnaires are appropriate to be implemented for this study. For example, the Sense of competence questionnaire includes items about statistics in general. However, the PS-I intervention is about calculating ranges or standard deviations. The questionnaire questions should be more detailed.

In the absence of measures of these predispositions that are validated and that can be more closely referred to the specific contents of this study, we used such instrument. This issue has been noted in the limitations section (lines 843-848).

Minor Concerns:
This manuscript needs to be edited and polished - Too many pages & Bulky sentences.

The manuscript has been reviewed by a native English reviewer (different corrections can be seen along the document).

RESPONSES TO THE REVIEW OF THE VIDEO

1. Composition

It looks like there's a warp stabilization filter on these clips. They warp and worble. Please remove this:
03:38, 03:40, 03:43, 05:29, 07:43, 07:54
These issues have been removed (now minutes 3:16, 3:18, 3:21, 5:07, 7:17, 7:28)

00:01 - 00:01 Please remove this logo.
The logo has been removed, both at the beginning and at the end of the video.

00:05 - 00:05 There's a watermark throughout the video. Please remove.
The watermark has been removed.

03:57 - 03:57 There's a weird letterboxing going on during this transition. Please remove the letterbox.
The letterbox has been removed (3:35).

03:58 - 03:58 The list appears on the left when it should appear in the middle. Please reposition this list so it's centered and slightly larger.

The list has been centered and enlarged to give a more organized idea (3:33)

04:25 - 04:25 Please center this figure.

The figure has been reorganized to appear in the center (3:56)

06:08 - 06:08 Please fade to white, not grey.

This fade (and others marked later) has been changed to white (5:43)

08:25 - 08:25 This figure can be enlarged and centered.

The figure has been enlarged and centered (8:05).

08:58 - 08:58 Please make this figure smaller so there's more room around the edge of the frame.

The figure has been removed according to reviews in the manuscript (8:31).

09:38 - 09:38 Please fade to white

This part has been retaped (according to a posterior suggestion) and faded to white (9:08)

2. Text

00:06 - 00:06 Please center the text from top to bottom.
The title of this image has been moved towards the bottom, here and at the end of the video (0:02)

03:01 - 03:01 I recommend putting ""about Ethical"" on the same line as ""Procedures"" so the top line and the bottom line are more even.

This title is more even now that has been arranged as suggested (2:42)

03:47 - 03:47 Place ""Measures"" on the bottom line with ""about Predipositions"" to make the top and bottom lines more even.

This title is more even now that has been arranged as suggested (3:23)

06:57 - 06:57 Place ""of the"" on the second line with ""Lecture"" so that the top and bottom lines are even.
This title is more even now that has been arranged as suggested (6:32)

07:12 - 07:12 Place ""Dependent"" on the bottom line with ""Variables""

This title is more even now that has been arranged as suggested (6:47)

08:05 - 08:05 The text on the top is too close to the edge of the frame. Please allow more space around the border. The figure also needs to be smaller.

The figure has been reduced to leave more margins (7:49)

10:43 - 10:43 Please center top to bottom.

The title of this image has been moved towards the bottom (10:11)

3. Pacing

I noticed there are no transitions between titles and video. Please fade out to white, then fade in to the text/video. Everytime a title comes up, the video before should fade to white, then once the screen is fully white, the title fades up, then the title fades to white, then the video fades up. I've marked all the time marks where this should occur:
00:13, 00:35, 00:54, 01:09, 03:00, 03:14, 03:47, 03:50, 04:15, 04:19, 05:47, 06:33, 06:55, 06:59, 07:11, 07:14, 07:31, 07:34, 08:00, 08:03, 08:23, 08:38, 08:56, 09:13, 09:35

All of these fades have been done as suggested (00:13, 00:34, 00:52, 1:08, 2:35, 2:50, 3:23, 3:25, 3:51, 3:54, 5:24, 6:08, 6:30, 6:34, 6:46, 6:49, 7:06, 7:09, 7:35, 7:38, 7:58, 8:14, 8:32, 9:05. Note that transition originally marked in minute 8:56 does not apply because they image has been removed).

01:34 - 01:34 There's a pause then a swallow. If there's a way to remove this by cutting out the pause and covering the jump cut with footage of the protocol.

This part has been videotaped again to remove this swallow and other problems commented later (1:33).

01:42 - 01:42 This dip to black and the repositioning of the talent is very very jarring. I had mentioned in the previous note to use footage from the protocol to cover up any jump cuts. I highly recommend you do that so this awkward cut can be avoided.

This cut has been removed by adding the suggested at this point (1:40)

01:43 - 01:43 It seems like at points he's unsure what he's going to say next. He's also constantly swallowing. I recommend re-recording this interview after he drinks some water and maybe write out a script so he has a more concise message.
The whole interview has been recorded.

01:52 - 01:52 Please start this figure at 1 minute and 34 seconds in, so it covers up the person pausing to swallow as well as the awkward dip to black.

The interview has been videotaped again, but we have used the figure anyway as suggested (1:40)

04:24 - 04:24 Instead of cutting in the rest of the figure, let's do a fade in.

The fade in has been made to make the change more smoother (4:02).

04:32 - 04:32 Instead of cutting in the rest of the figure, please fade in.

The fade in has been made to make the change more smoother (4:10).

04:45 - 04:45 The image cuts to white, then fades to the next figure. Please fade out the video.

The video prior to the image has been faded out (4:19)

09:07 - 09:07 Please fade on these text and arrows.

This image has been removed according to reviews in the manuscript (8:30).

09:39 - 10:41 I can see the person keeps looking down at the script after every sentence. It's very distracting and should be refilmed without looking down at the script.

This part has been refilmed without looking at the script (9:06)

10:16 - 10:16 The person pauses to swallow. Please remove this and cover the rest of the interview with footage from the protocol.
This whole interview has been refilmed and this problem has been avoided (9:46)

10:44 - 11:06 The end title slides last 23 seconds. Please reduce the amount to 12 seconds. That will give about 4 seconds for each slide. Please remove the logo at the end."

This timeframe has been reduced as suggested, and the logo has been removed (10.:08-10:20)

4. Audio

02:55 - 02:55 The interview audio gets cut off too quickly. Please allow the audio to fade out with the video.

This audio transition has been corrected to fade at the same time than the video (2:33)

04:52 - 04:52 I can hear the narrator swallow. Please remove.

The swallow sound has been removed (4:29)

05:12 - 05:12 I can hear a page flip. Please remove any audio that's not scripted.

The page flip sound has been removed (4:49).

05:19 - 05:19 I can hear the narrators mouth opening. Please remove any space between the sentences so any background audio or unscripted audio can't be heard.

This sound has been removed (4:55)

06:31 - 06:31 I think I heard the videographer in the background talking. Please remove this.

This sound has been removed (6:06).

06:58 - 06:58 There's a blip of noise. Please remove.

This sound effect has been removed (6:35)

07:40 - 07:40 There's a weird dip in the audio that cuts off part of the narration. Please fix this.

This cut in the audio has been removed (7:18)

Click here to access/download;Rebuttal Letter;Response to video reviewers.docx

Manuscript ID: JoVE62138R1
Title: Problem-Solving Before Instruction (PS-I): A protocol for assessment and intervention in students with different abilities

Dear editors and reviewers,

We acknowledge all the comments, which have contributed to improve the quality of the video.

Below, we provide a detailed response to all the comments and suggestions.

1. Composition
03:32 There's a weird letterboxing going on during this transition. It looks like the figure is being cropped.

The weird letterboxing has been eliminated from the transition (3:32)

08:15 The Text and the figures are way too close to the edge of the frame. Please resize the text and adjust the figures so they aren't so close to the bottom.

The image has been reduced to leave more margins (8:15)
08:32 Please make this figure smaller so there's more room around the edge of the frame.

The image has been reduced to leave more margins (8:32)
2. Text
00:04 Please center the text from top to bottom. Right now it's too low in the frame. Please shift it up a bit.

It has been centered (0:01)

05:24 I recommend placing ""During the"" on the second line with ""Invention Problem"".

It has been reorganized as recommended (5:24)

07:08 I recommend either placing the title all on one line, or moving ""Analyses"" to the second line.

It has been reorganized moving “Analyses” to the second line (7:07)

10:08 Please center the text top to bottom.

It has been centered

3. Pacing

00:00 The video starts with two seconds of black. Please remove this and start the video with the fade up.

The video now starts directly with the fade up

01:38 Please fade out the person talking, instead of cutting out.

The fade out has been made (1:37)

01:59 Please fade out, then fade in.

This transition has been done as recommended (1:58)

10:20 There's an extra three seconds of black at the end. Please remove this.

It has been removed (10:17)

4. Audio
06:06 I can hear a hiss where the videographer was speaking at the end. It sounds like you covered the audio with another source of audio not from this interview. I recommend using a piece of this interviews audio to cover up the videographers voice.

This part of the audio has been covered as recommended (6:06-6:08)

07:15 There's still a weird dip in the audio. The audio clips need to fade between each other, not abruptly stop then start.

This was occurring because a previous edition cutting the audio was needed and is hard to modify, but the audiovisual Department has tried to dissimulate this abrupt effect (7:15)

Click here to access/download;Rebuttal Letter;JoVE62138_R2 Response to reviewers.docx

Submission ID: JoVE62138R2
Title: Problem-Solving Before Instruction (PS-I): A protocol for assessment and intervention in students with different abilities

Dear editors and reviewers,

We acknowledge all the comments, which have contributed to keep improving the quality of the video. However, we must recognize that there is one suggestion that we have not been able to completely address, concerning the audio.

Below, we provide a detailed response to all the comments and suggestions.

1. Composition
03:32 - There's a weird letterboxing going on during this transition. Please resize the figure so that the white background completly fills the screen.

The Figure has been resized to eliminate the weird letterboxing.

2. Text
00:04, 10:08 - Plesae center the text from top to bottom by bringing it up a bit. It's currently too low, for both the intro and outro.

The image has been centered.

3. Audio
07:15 - There's still a weird dip in the audio. Please remove the fade out on the word "protocol".

This aspect could not be totally solved in the current version of the video. As the Audio-visual Department of our University informed us, they have been working in removing the fade out and the dip, and it has been substantially improved. However, they have not been able to completely solve the issue, because this artifact was originated as a result of a previous change made on the video (it was previously transformed to change a pronunciation error).

Click here to access/download;Rebuttal Letter;JoVE62138_R3 Response to reviewers.docx

Submission ID: JoVE62138R3
Title: Problem-Solving Before Instruction (PS-I): A protocol for assessment and intervention in students with different abilities

Dear editors and reviewers,

We acknowledge all the comments, which have contributed to keep improving the quality of the video.

Below, we provide a detailed response to all the comments and suggestions.

Please add an ethics card before the protocol section of the video.

The card has been added (min: 2:36)

07:15 - The fade was removed, but now there's a blip that you can hear. I highly recommend re-recording this part of the narration so it doesn't sound like it's pieced together with different takes.

The audio of the narrator has been re-recorded to solve this issue (min 7:18)

Supplemental File (Figures, Permissions, etc.)
Click here to access/download;Supplemental File (Figures, Permissions, etc.);SUPLEMENTARY MATERIALS.docx

Supplementary Material

Contents

Appendix A. Informed Consent (Step 1)

Informed Consent Form

The present study follows the Helsinki Declaration of Ethical Principles for the Investigation with Humans and has been approved by the Ethical Committee for the Investigation of the Principality of Asturias (Reference: 242/19).

Objectives and description of the study. The first objective is to explore the efficacy of two educational approaches for learning statistics. To facilitate this comparison, you and your classmates will be randomly assigned to perform different activities. Yet, all of you would cover the same contents. Also, activities of the two conditions will be available to all of you at the end. The second objective is to identify the cognitive factors that can foster learning in both approaches. For this, you and your classmates will be asked to complete short questionnaires and tests before starting the learning process.

Free decision to participate. You are free to participate in the study or abandon it at any time without any prejudice to your learning opportunities nor to your grade. If you do not want to participate, you can perform the learning activities without handing them in, and during the short time that other students are answering the questionnaires, you would be allowed to study other materials.

Confidentiality and anonymity. Your responses cannot be associated with your identity at any time. To help us connect your responses from different activities and questionnaires, you will be assigned an arbitrary identification number, and you will be asked to report this number in each activity.

Feedback about learning. At the end of the evaluation process you will have access to the correct responses in the learning tests, together with all the learning activities used in all conditions.

Information about results. Once the data is analyzed, your professor will give you access to the results of the investigation.

Further information. If you have any questions related to this investigation you can contact: (to be completed with data of the principal researcher)**.**

After having read the information in this document and to consent to your participation in the present investigation, please sign below. You are allowed to interrupt your participation at any time, for any reason.

Signed: Place: Date:

Appendix B. Demographic and Predispositions Questionnaires (Step 3)

1. Printed Versions of the Three Questionnaires Together with Four Counterbalanced Orders

The four printed versions, each with a different order presentation, can be accessed at the following link: https://www.dropbox.com/sh/uv9t85wabivvurm/AAAVSezPwdcJuU6cf4gZSvdxa?dl=0

1. Demographic Information Questionnaire

Age: Sex: Male/Female

Course: Educational Institution:

2. Sense of Competence Questionnaire

Please describe your level of agreement regarding the following statements about how you feel in learning statistics

	Strongly Disagree			Neither agree nor disagree			Strongly Agree
1. I will have trouble understanding statistics because of how I think	1	2	3	4	5	6	7
2. I will have no idea of what is going on in this statistics course	1	2	3	4	5	6	7
3. I will make a lot of math errors in statistics	1	2	3	4	5	6	7
4. I can learn statistics	1	2	3	4	5	6	7
5. I will understand statistics equations	1	2	3	4	5	6	7
6. I will find it difficult to understand statistical concepts	1	2	3	4	5	6	7

3. Mastery Approach Goals Questionnaire

Please describe your level of agreement of these statements about your goals in this class

	Totally Disagree		Neither agree nor disagree		Totally Agree
1. My aim is to completely master the material presented in this class	1	2	3	4	5
2. I am striving to understand the content of this course as thoroughly as possible	1	2	3	4	5
3. My goal is to learn as much as possible	1	2	3	4	5

4. Metacognitive Regulation Questionnaire

The purpose of this questionnaire is assessing your awareness about learning. For each statement, please respond how typical they are on you

	Not at all typical of me	**Not very** typical of me	**Somewhat** typical of me	**Fairly** typical of me	**Very** typical of me
1.I ask myself periodically if I am meeting my goals	1	2	3	4	5
2. I ask myself if I have considered all options when solving a problem	1	2	3	4	5
3. I consciously focus my attention on important information	1	2	3	4	5
4. I ask myself if there was an easier way to do things after I finish a task	1	2	3	4	5
5. I consider several alternatives to a problem before I answer	1	2	3	4	5
6. I periodically review to help me understand important relationships	1	2	3	4	5
7. I ask myself questions about the material before I begin	1	2	3	4	5
8. I think of several ways to solve a problem and choose the best one	1	2	3	4	5
9. I summarize what I've learned after I finish	1	2	3	4	5
10. I ask others for help when I don't understand something	1	2	3	4	5
11. I find myself analyzing the usefulness of strategies while I study	1	2	3	4	5
12. I focus on the meaning and significance of new information	1	2	3	4	5
13. I create my own examples to make information more meaningful	1	2	3	4	5
14.I find myself pausing regularly to check my comprehension	1	2	3	4	5
15. I ask myself how well I accomplish my goals once I'm finished	1	2	3	4	5
16. I draw pictures or diagrams to help me understand while learning	1	2	3	4	5
17. I ask myself if I have considered all options after I solve a problem	1	2	3	4	5
18. I try to translate new information into my own words	1	2	3	4	5
19. I pace myself while learning in order to have enough time	1	2	3	4	5
20. I change strategies when I fail to understand	1	2	3	4	5
21. I use the organizational structure of the text to help me learn	1	2	3	4	5
22. I read instructions carefully before I begin a task	1	2	3	4	5
23. I ask myself if what I'm reading is related to what I already know	1	2	3	4	5
24. I reevaluate my assumptions when I get confused	1	2	3	4	5
25. I organize my time to best accomplish my goals	1	2	3	4	5
26. I try to break studying down into smaller steps	1	2	3	4	5
27. I focus on overall meaning rather than specifics	1	2	3	4	5
28. I ask myself questions about how well I am doing while I am learning something new	1	2	3	4	5
29. I ask myself if I learned as much as I could have once I finish a task	1	2	3	4	5
30. I stop and go back over new information that is not clear	1	2	3	4	5
31. I stop and reread when I get confused	1	2	3	4	5
32. I think about what I really need to learn before I begin a task	1	2	3	4	5
33. I know how well I did once I finish a test	1	2	3	4	5
34. I set specific goals before I begin a task	1	2	3	4	5
35. I slow down when I encounter important information	1	2	3	4	5

Appendix C: Pre-test of Previous Academic Knowledge (Step 5)

Identification number:

Please answer the following questions:

1) The following shows the timing expressed in seconds for a group of 9 students in a 25 meters swimming competition: 12, 14, 16, 14, 18, 16, 13, 14, 18
¿What is the mean, the median, and the mode of the timing of this group?

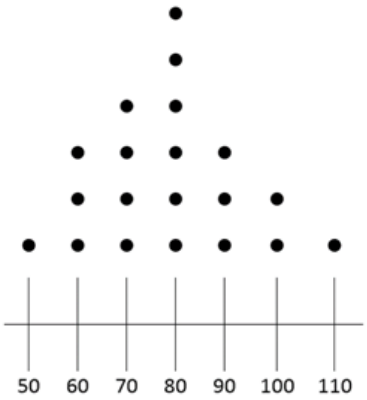
2) In the previous distribution, ¿What would the quartile 1 and the quartile 3 be?

3) In a different group of 5 students, the mean for the 25 meters swimming competition is 20 seconds. The incorporation of an additional member to this group changes the group mean to 19 seconds. What is the time of this additional student?
   a. 119 seconds
   b. 19 seconds
   c. 14 seconds
   d. 15 seconds
   e. None of the above

4) What is the most appropriate measure of central tendency to address the performance of this group of students? Why?
   a. La median, because it is an ordinal variable.
   b. The mean, because it is an ordinal variable.
   c. The median, because it is an interval variable.
   d. The mean, because it is a ratio variable.
   e. None of the above.

1) The heart rate per minute of a group of 20 adults is displayed in the dot diagram below. For example, 3 adults have a rate of 60 beats per minute. Based on this data set, how many individuals from a similar group of 40 adults would be expected to have a heart rate of at least 90 beats per minute?

   a. **1**2 adults
   b. 6 adults
   c. 3 adults
   d. 28 adults
   e. None of the above



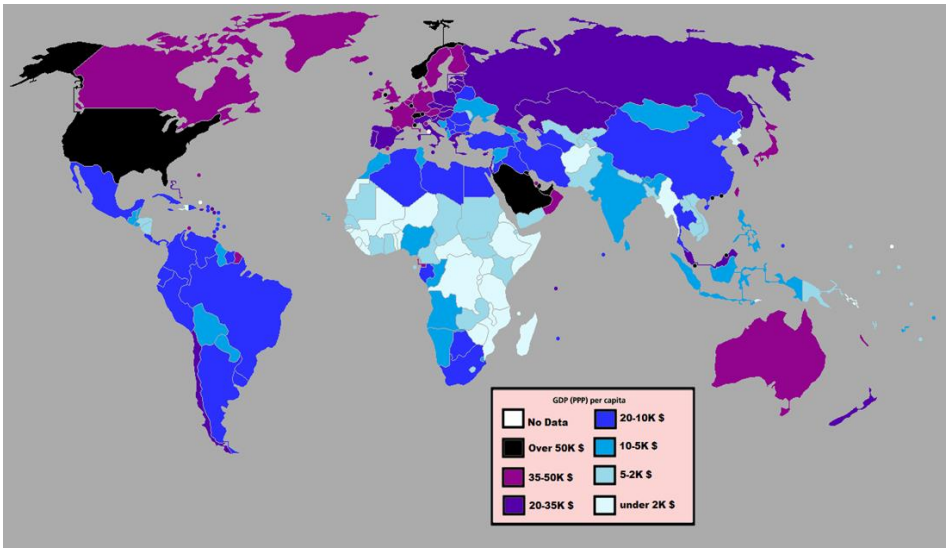
Appendix D: Materials of the Two Learning Conditions (Step 6)

1. Printed Version of all Materials Together for the Two Conditions

The taskbook for each learning conditions, the PS-I condition and the direct instruction condition, can be accessed at the following link:

https://www.dropbox.com/sh/wy52p6wvxgkebcn/AADWPjSby5ZszYBeCyyBFFZVa?dl=0

2. Worked Example



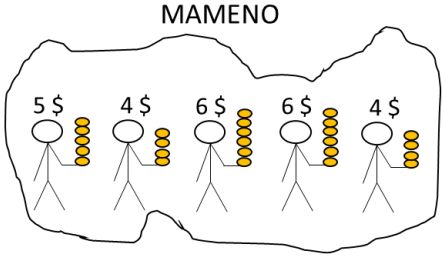
This image shows the average income (GDP per capita) of the countries in 2012. We can see that, for example, Spain and Russia have a similar average income. However, only with the average income, can we have an idea of the wealth of the inhabitants?

Below you can see four imaginary countries. Although the average income is the same in all of them, it is easy to see that they differ on how inequality strongly affects the wealth of their inhabitants.
How can we measure inequality?

There are multiple solutions that can help us to measure inequality and compare the inequality of these four countries. In the following pages we will show you three of the most accepted solutions.

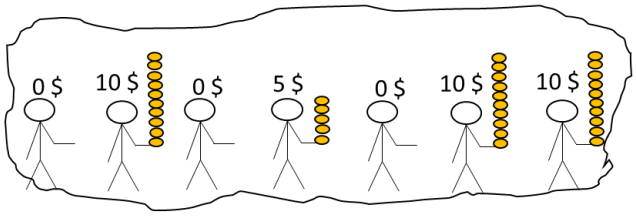
Try to learn their calculation procedures and comprehend their affordances and limitations.

MAMENO



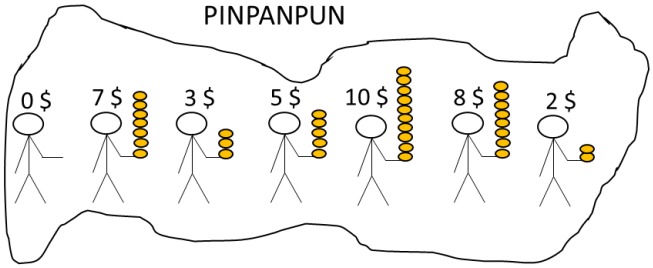
Average Income: (5+ 4 + 6 + 6 + 5) / 5 = **5**

TOVEO



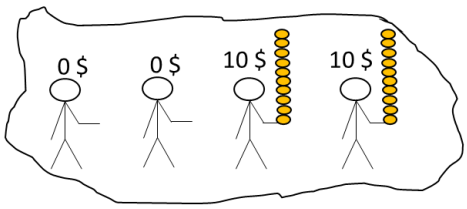
Average Income: (0 + 10 + 0 + 5 + 0 +10 + 10) / 7 = **5**

PINPANPUN



Average Income: (0 + 7 + 3 + 5 + 10 + 8 + 2) / 7 = **5**

MITAPAI



Average Income: (0 + 0 + 10 + 10) / 4 = **5**

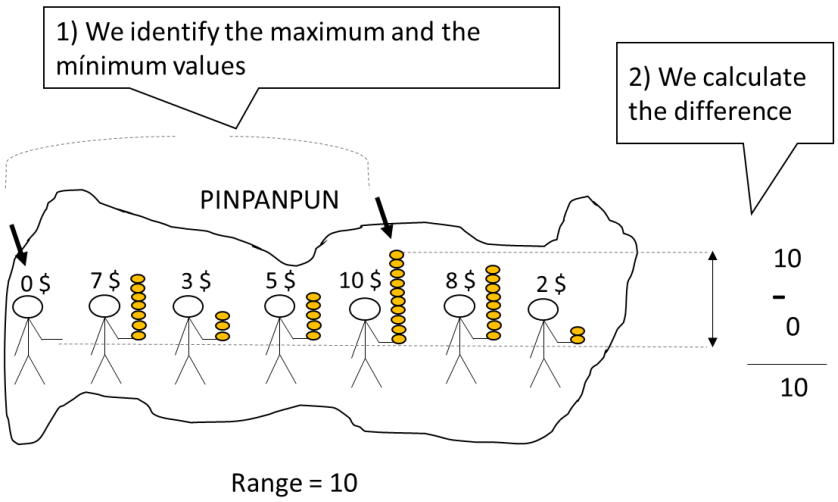
(continue to the next page)

SOLUTION 1: THE RANGE

- It is the difference between the maximum and the minimum value of the distribution.

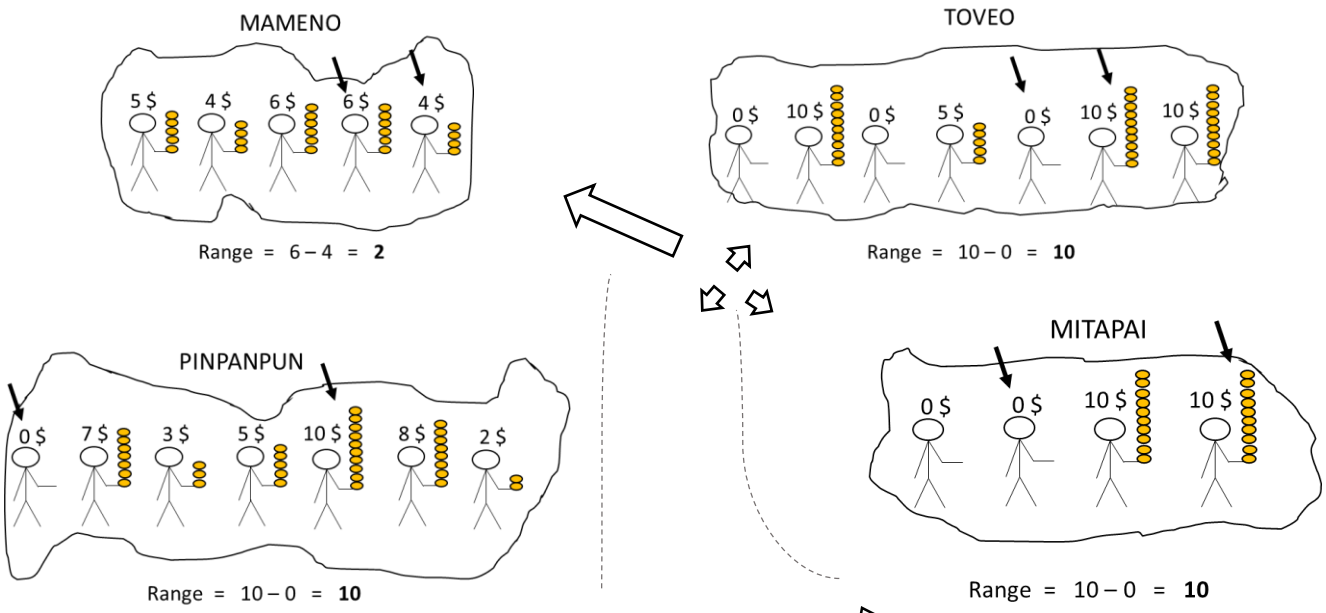
How is it calculated?

$$R = \max_i(x_i) - \min_i(x_i)$$



Range = 10

Solution:



ADVANTAGES OF THE RANGE:

It is easy to be calculated, and in some cases is a good indicator of inequality. For example, the lower range of Mameno certainly corresponds with a lower inequality in this country than in the rest of the countries.

LIMITATIONS:

It only takes into account the two extreme values of the distribution, the richest and the poorest person. Therefore, it is not possible to identify the different degree of inequality between Toveo, Mitapai and Pinpanpun, in which the extreme values are the same.

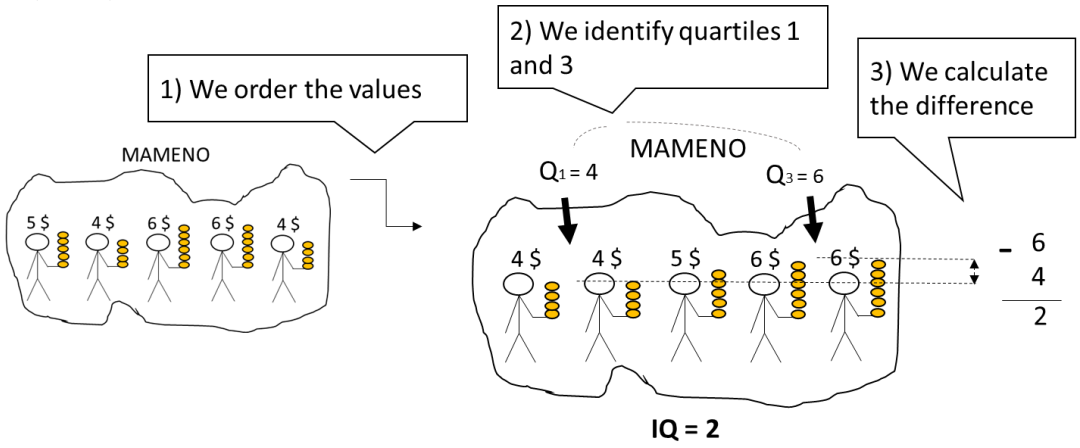
(continue to the next page)

SOLUTION 2: INTERQUARTILE RANGE (IQ)

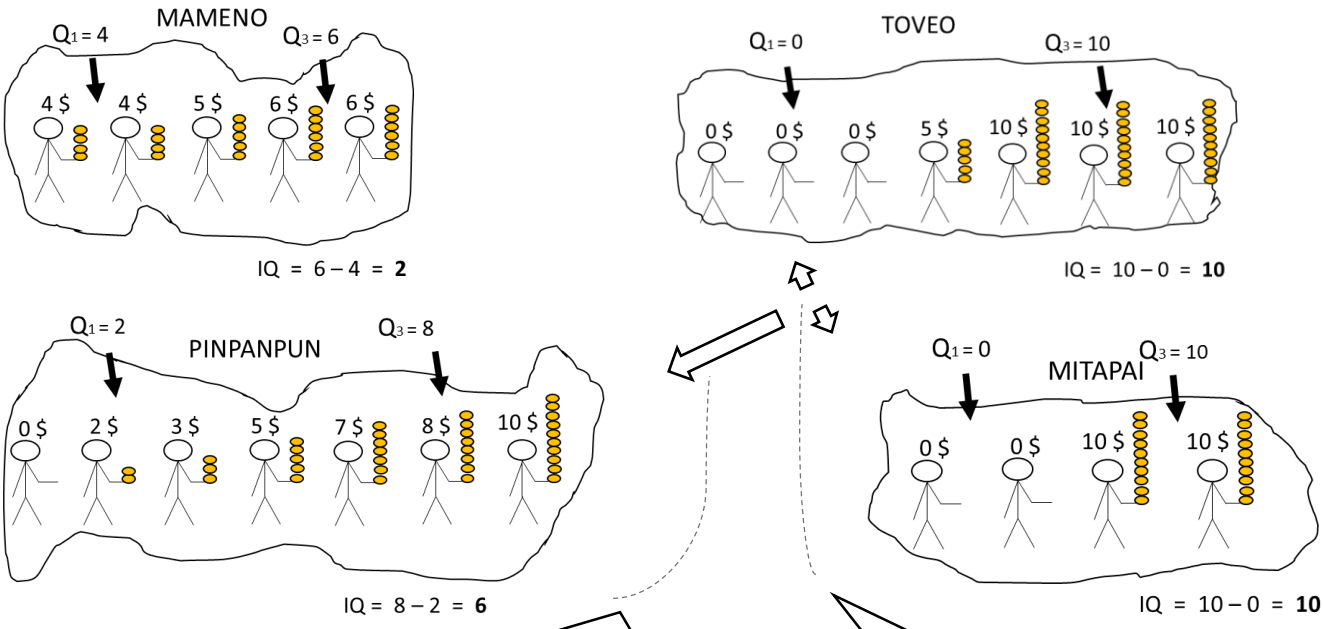
- It is the difference between quartile 3 and quartile 1. In other words, the amplitude of the range that contains the central half of the distribution.

How is it calculated?

$IQ = Q_3 - Q_1$



Solution:



ADVANTAGES OF THE INTERQUARTILE RANGE:

It is not distorted by extreme scores because it only considers the central differences.

It is relatively comprehensive. Although it only considers differences in the central half of the distribution, these differences are expected to be representative of differences in the whole distribution. It allows us to see that there is more equality in Pinpanpun (IQ = 6) than in Toveo and Mitapai (IQ = 10).

LIMITATIONS OF THE INTERQUARTILE RANGE:

Its exhaustiveness can be improved because it does not consider the differences across all cases. That is why we cannot identify the different levels of inequality between Toveo and Mitapai, where Q_1 and Q_3 are the same.

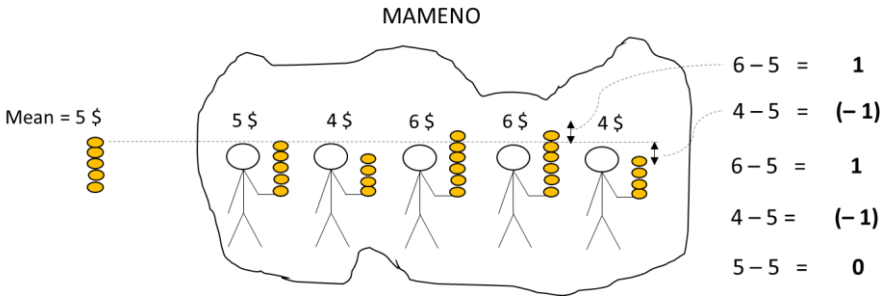
(continue to the next page)

SOLUTION 3: THE STANDARD DEVIATION

- It is an average of the differences from the mean.

How is it calculated?

1) We calculate the differences from the mean $x_i - \overline{x}$



2) We squared them. $(x_i - \overline{x})^2$

This step is necessary to convert the differences in possitive values. Otherwise, when summing them up in the next step the negative values would supress the positive values. Ej:

1 + (-1) + 1 + (-1) + 0 = 0

$1^2 = 1$

$(-1)^2 = 1$

$1^2 = 1$

$(-1)^2 = 1$

$0^2 = 0$

3) We average the differences: We first sum them up, and then we divide the total by the number of cases. With it we obtain the variance (σ^2)

$$\frac{\Sigma (x_i - \overline{x})^2}{N}$$

$$\frac{1+1+1+1+0}{5} = \frac{4}{5} = 0{,}8$$

$\sigma^2 = 0{,}8$

4) Because before we have squared the differences, we have to invert this process calculating the square root. With it we obtain the standard deviation (σ).

$$\sqrt{\frac{\Sigma (x_i - \overline{x})^2}{N}}$$

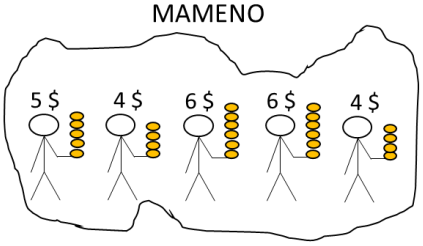
$\sqrt{0{,}8} = 0{,}89$

$\sigma = 0{,}89$

(continue to the next page)

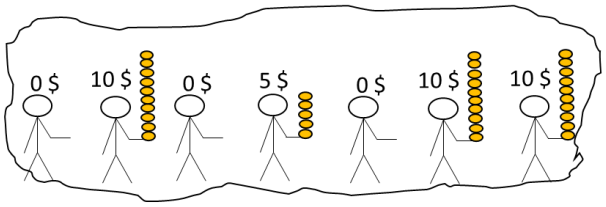
Solution:

MAMENO



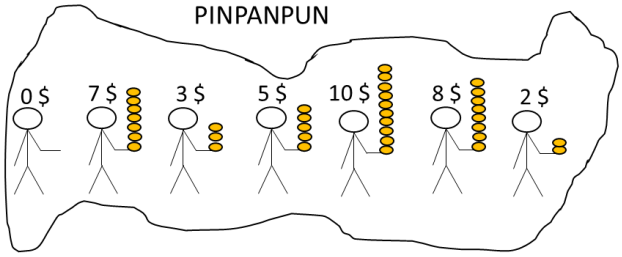
$\sigma = \sqrt{\frac{(5-5)^2+(4-5)^2+(6-5)^2+(6-5)^2+(4-5)^2}{5}} = \mathbf{0,89}$

TOVEO



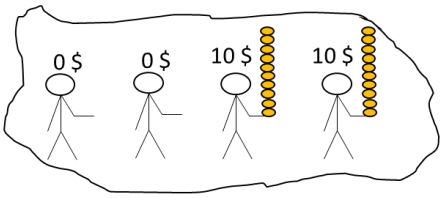
$\sigma = \sqrt{\frac{(0-5)^2+(10-5)^2+(0-5)^2+(5-5)^2+(0-5)^2+(10-5)^2+(10-5)^2}{7}} = \mathbf{4,62}$

PINPANPUN



$\sigma = \sqrt{\frac{(0-5)^2+(7-5)^2+(3-5)^2+(5-5)^2+(10-5)^2+(8-5)^2+(2-5)^2}{7}} = \mathbf{3,29}$

MITAPAI



$\sigma = \sqrt{\frac{(0-5)^2+(0-5)^2+(10-5)^2+(10-5)^2}{4}} = \mathbf{5}$

ADVANTAGES OF THE STANDARD DEVIATION:

It is a very comprehensive indicator, as it considers the differences of all the cases in the distribution. This allow us to identify the different degrees of inequality of the 4 countries, even when they are very subtle, for example when we compare Mitapai and Toveo.

LIMITATIONS OF THE STANDARD DEVIATION:

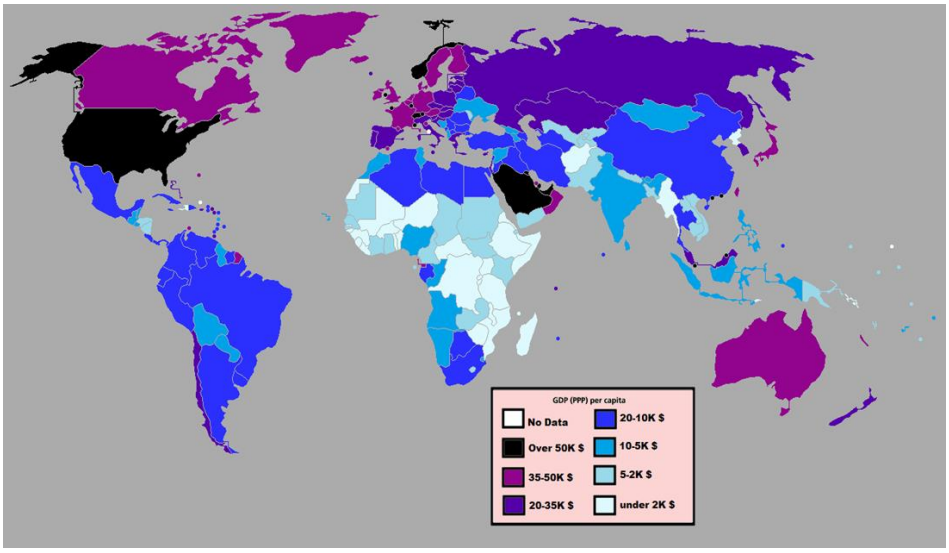
- Although in general it is quite resistant to extreme values, it also considers them in its calculation. This is the reason why, when we find very asymmetric distributions, where there are very extreme values towards one side of the distribution, the standard deviation can be distorted.

- It is not always appropriate to be used with ordinal variables. In ordinal variables the differences between values are uncertain, and the standard deviation is based on the calculation of these exact differences.

When we are in any of these situations, we can use the interquartile range.

Only if you have finished studying these contents, in the reverse of the page you can practice developing the procedures for the calculation of the inter-quartile range and standard deviation in Toveo , Pinpanpun, and Mitapai. Do your solutions correspond with the solutions given here?

3. Invention Problem

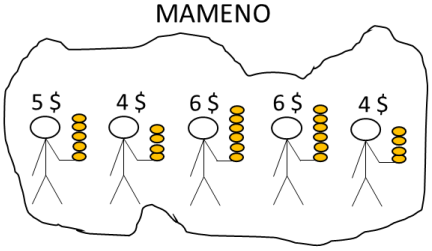


This image shows the average income (GDP per capita) of the countries in 2012. We can see that, for example, Spain and Russia have a similar average income. However, only with the average income, can we have an idea of the wealth of the inhabitants?

Below you can see 4 imaginary countries. Although the average income is the same in all of them, it is easy to see that they differ on how inequality strongly affects the wealth of their inhabitants.
How can we measure inequality?

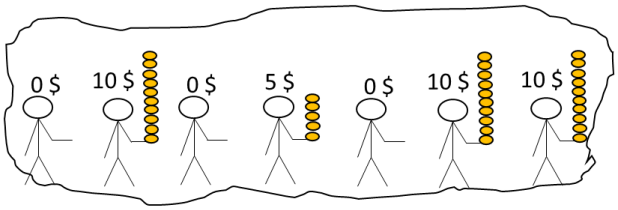
Try to design a mathematical index that would help us to measure and compare inequality in these 4 countries. You can give multiple answers. There is no one unique valid solution!

MAMENO



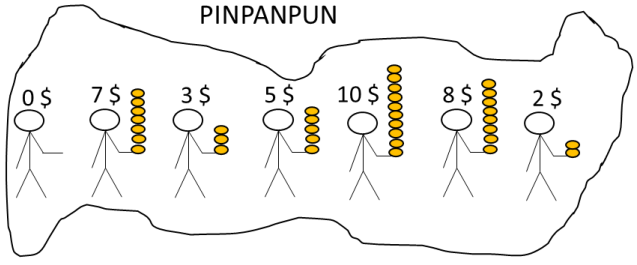
Average Income: (5+ 4 + 6 + 6 + 5) / 5 = **5**

TOVEO



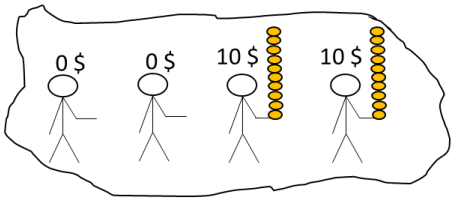
Average Income: (0 + 10 + 0 + 5 + 0 +10 + 10) / 7 = **5**

PINPANPUN



Average Income: (0 + 7 + 3 + 5 + 10 + 8 + 2) / 7 = **5**

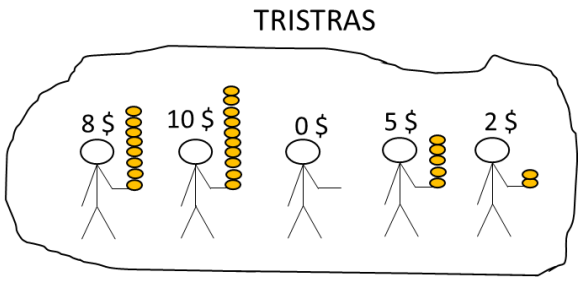
MITAPAI



Average Income: (0 + 0 + 10 + 10) / 4 = **5**

4. Practice Problem

Calculate the range, the interquartile range, and the standard deviation for the income distribution in Tristras. Discuss the affordances and limitations of how these three measures allow to observe higher or lower inequality in Tristras versus the four countries you have seen in the previous activity.



Appendix E: Further description of the Lecture (Step 7)

The power-point to support the lecture can be seen in https://www.dropbox.com/sh/aa6p3hs8esyf5xa/AACTvpVlEbdEtLVfBIbe9j7aa?dl=0. This file includes animations to stagger the contents, comments with proposed explanations to give to students, and indications about the approximative time allocated for each explanation. Also, a general description of these aspects is provided below:

1. Introductory explanation of the concept of variability, the different types of variability measures, their complementation with central tendency measures, and the use of all of them depending on the different type of variables. Slides 1-5. Approximate time: 3 minutes.

2. Introductory explanation of the standard deviation, when we use it and why. Slides 6-8. Approximate time: 4 minutes.

3. The students are given time to calculate the standard deviation for the distributions in the following problem. They can refer back to the explanations in the Worked Example. Slide 9. Approximate time: 5 minutes.

   PROBLEM: The scores obtained by two groups of 5 students in a validated test of knowledge are:

   - Group A: 4, 5, 6, 7, 8
   - Group B: 4, 6, 6, 6, 8

   In which group of students is there more variability? Justify your answer with the calculation of a variability measure

4. Feedback on the calculations and interpretation of results in the previous problem. Slides 10-11. Approximate time: 2 minutes.
5. Feedback on incorrect or partial calculations of the standard deviation. Slides 12-21. Approximate time: 7 minutes**.**
6. Short explanation of the different symbols used to refer to standard deviation depending on whether we refer to samples or populations. Slide 22. Approximate time: 1 minute.
7. Short explanation of the formula used in SPSS or other inferential statistics programs to calculate the standard deviation: the unbiased standard deviation. Slide 23. 1 minute.
8. Explanation of the properties of the standard deviation. Slides 24-25. Approximate time: 2 minutes.
9. Introductory explanation of the interquartile range: when and why we use it. Slides 26-30. Approximate time: 5 minutes.
10. Students are given time to calculate the interquartile range of the distributions in the following problem. They can refer back to the explanations in the Worked Example. Slide 31. Approximate time: 3 minutes.

PROBLEM: The scores obtained by 2 groups of 7 students in a exam are:

GRUPO A	GRUPO B
Outstanding (3)	Outstanding (3)
Failed (0)	Failed (0)
Very Good (2)	Outstanding (3)
Honors (4)	Very Good (2)
Very Good (2)	Failed (0)
Aprobado (1)	Failed (0)
Very Good (2)	Outstanding (3)

¿In which group of students there is more variability? Justify your answer with the calculation of a variability measure

11. Feedback on the calculations and interpretation of the interquartile range from the previous problem, including the discussion of the partial solution of the range. Slides 32-33. Approximate time: 2 minutes.
12. Explanation of the property of resistance to extreme values and the appropriateness of the interquartile range when we deal with ratio or interval variables in asymmetric distributions. Slide 34. Approximate time: 2 minutes.
13. Introductory Explanation to the Coefficient of Variation: How we calculate it, when and why we use it. Slides 35-37. Approximate time: 3 minutes.
14. Overview of the contents covered. Slide 38. Approximate time: 1 minute

Appendix F: Curiosity Questionnaire (Step 8)

Identification number:

We are interested in the emotions you experienced when learning this lesson. For each emotion, please indicate the strength by circling the number that best describes the intensity of your emotional response during the learning process.

	Not at all	Very little	Moderate	Strong	Very Strong
1.Curious	1	2	3	4	5
2. Interested	1	2	3	4	5
3. Inquisitive	1	2	3	4	5

Appendix G: Learning Post-test (Step 9)

Identification number:

Please answer the following questions:

1. (a) Calculate the standard deviation (SD) of the following set of marks on a test:

   3, 6, 5, 4, 7

   A. 1 B. $\sqrt{2}$ C. 2 D. $\sqrt{1}$

2. The owners of two cinemas, A and B, argue that their respective cinema enjoys a more consistent attendance. They collected the daily attendance of their cinemas for 11 random days. The results of their data collection are shown below:

	Cinema A	Cinema B
Mean, M	*72*	*75*
Standard Deviation, SD	*10*	*14*

Which cinema do you think presents a more consistent attendance?

A. Cinema A
B. Cinema B
C. Both enjoy equally consistent attendance.
D. None of the above.

3. Below we can see the data from 2 distributions. In which distribution there is more variability?

	Distribution A	Distribution B
Median	*6*	*8*
Interquartile Range	*3*	*2*

A. In distribution A
B. In distribution B
C. Both have the same variability
D. Not enough information to decide

4. In calculating the standard deviation, why is it important to divide the sum of squared deviations by *n*?

5. As we can see in the following table, the standard deviation is greater in group A than in group B, while the interquartile range is greater in group B than in group A. How can it be explained?

	Group A	Group B
Standard Deviation	5,37	2,77
Interquartile Range	2	4

   a) In the Group A there must be more extreme scores than in Group B.
   b) In the Group B there must be more extreme scores than in Group A.
   c) The interquartile range indicates the central tendency, and the standard deviation indicates the variability, and therefore they do not have to coincide.
   d) It is imposible, both are measures of variability and they have to provide consistent results.

6. Consider the following six datasets:
   A) (1, 5, 6, 10)
   B) (4, 4, 4, 4)
   C) (101, 102, 103, 104)
   D) (7, 8, 9, 10)
   E) (1, 2, 9, 10)
   F) (1, 2, 3, 4)

   a) Which dataset has the smallest SD? _____
   b) Which dataset has the largest SD? _____
   c) Which datasets have the same SD? _____

7. A data set consisting of <u>five</u> numbers has mean, M = 7, and standard deviation, SD = 4. If each of the five numbers is increased by 2, what are the new mean and SD?
   A. M = 7, SD = 4
   B. M = 9, SD = 4
   C. M = 7, SD = 6
   D. M = 9, SD = 6

8. An equal number of students competed in the 100m sprint and 100m swim finals. The timings (in seconds) of the champions of the 100m sprint and 100m swim are shown below, as are the average timings and the SDs of the finalists in the two competitions.

	100m sprint	100m swim
Champion	10s	40s
Average of the Finalists, M	12s	45s
SD of the Finalists	1s	3s

Assuming all else being equal, between the two champions, who is the better performer?

A. The sprint champion
B. The swim champion
C. Both
D. Not enough information to decide

9. David's scores for Mathematics, Physics and Chemistry in the final examinations are given below. His class's performance for the three subjects is also given below:

	Mathematics	Physics	Chemistry
David's Score	95	90	85
Class Average, M	80	80	80
Class SD	15	5	4

a) Relative to his class, in which subject did David perform the best? _____
b) Relative to his class, in which subject did David perform the worst? _____

Appendix H. Materials to be Provided to Students at the End (Step 10)

A folder with the materials used during the lesson and the feedback for exercises in the pretest and the posttest can be downloaded in the following link:

https://www.dropbox.com/sh/88k9b8qgxngbukq/AABMIIxNc20jGGwhCTVal-F1a?dl=0

Concretely, the folder includes four documents:

- The “Solutions to Central Tendency Exercises” document includes the solutions to the pre-test exercises.
- The “Preparatory Exercises for Variability” document includes in a single document the Invention Problem, the Worked Example, and the Practice Problem.
- The “Variability Lecture Slides” document include the slides used for the lecture, but without the comments about the proposed explanations
- The “Solutions to Variability Exercises” document includes the solutions to the posttest exercises.

Appendix I. Answer Keys for Coding (Step 11)

1. Pretest Answer Key

The score of this pretest can vary from 0 to 8 points. It will be calculated summing the points obtained in each of the questions, as described below.

1) The following shows the timing expressed in seconds for a group of 9 students in a 25 meters swimming competition: 12, 14, 16, 14, 18, 16, 13, 14, 18
¿What is the mean, the median, and the mode of the timing of this group?
Mean: 15 → 1 point
Median: 14 → 1 point
Mode: 14 → 1 point

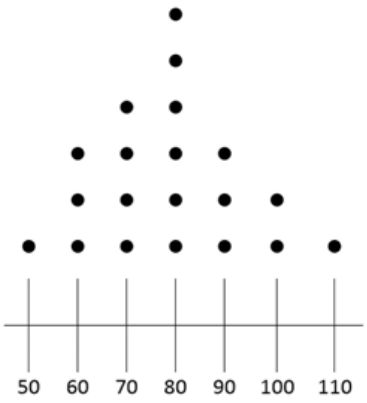
2) In the previous distribution, ¿What would the quartile 1 and the quartile 3 be?
Quartile 1: 13.5 → 1 point
Quartile 2: 17 → 1 point

3) In a different group of 5 students, the mean for the 25 meters swimming competition is 20 seconds. The incorporation of an additional member to this group changes the group mean to 19 seconds. What is the time of this additional student?
   a. 119 seconds
   b. 19 seconds
   c. 14 seconds → 1 point
   d. 15 seconds
   e. None of the above

4) What is the most appropriate measure of central tendency to address the performance of this group of students? Why?
   a. La median, because it is an ordinal variable.
   b. The mean, because it is an ordinal variable.
   c. The median, because it is an interval variable.
   d. The mean, because it is a ratio variable. → 1 point
   e. None of the above.

2) The heart rate per minute of a group of 20 adults is displayed in the dot diagram below. For example, 3 adults have a rate of 60 beats per minute. Based on this data set, how many individuals from a similar group of 40 adults would be expected to have a heart rate of at least 90 beats per minute?

   a. 12 adults → 1 point
   b. 6 adults
   c. 3 adults
   d. 28 adults
   e. None of the above



2. Posttest Answer Key

The total score of this posttest can vary from 0 to 12 points, and it will constitute the general learning. This measure can be disaggregated in other learning measures:

- The procedural knowledge score can vary from 0 to 3, and it will be calculated summing up scores in items 1-3.

- The conceptual knowledge score can vary from 0 to 6, and it will be calculated summing up scores in items 4-7.

- The transfer of knowledge score can vary from 0 to 3, and it will be calculated summing up scores in items 8-9.

The assignment of scores for each item is described below.

1. (a) Calculate the standard deviation (SD) of the following set of marks on a test:
   3, 6, 5, 4, 7

   A. 1 B. $\sqrt{2}$ → 1 point C. 2 D. $\sqrt{1}$

2. The owners of two cinemas, A and B, argue that their respective cinema enjoys a more consistent attendance. They collected the daily attendance of their cinemas for 11 random days. The results of their data collection are shown below:

	Cinema A	Cinema B
Mean, M	*72*	*75*
Standard Deviation, SD	*10*	*14*

Which cinema do you think presents a more consistent attendance?

   A. Cinema A → 1 point
   B. Cinema B
   C. Both enjoy equally consistent attendance.
   D. None of the above.

3. Below we can see the data from 2 distributions. In which distribution there is more variability?

	Distribution A	Distribution B
Median	*6*	*8*
Interquartile Range	*3*	*2*

   A. In distribution A → 1 point
   B. In distribution B
   C. Both have the same variability
   D. Not enough information to decide

4. In calculating the standard deviation, why is it important to divide the sum of squared deviations by *n*?

Assign 1 point if mentioned the need to average differences, or made reference to the idea that otherwise the indicator would be contaminated by the number of cases.

Assign 0.5 points if just mentioned a general need for averaging

Assign 0 points if none of the above.

Double coding of this item and evaluation of interrater reliability is recommended

10. As we can see in the following table, the standard deviation is greater in group A than in group B, while the interquartile range is greater in group B than in group A. How can it be explained?

	Group A	Group B
Standard Deviation	5,37	2,77
Interquartile Range	2	4

a) In the Group A there must be more extreme scores than in Group B.
b) In the Group B there must be more extreme scores than in Group A. → 1 point
c) The interquartile range indicates the central tendency, and the standard deviation indicates the variability, and therefore they do not have to coincide.
d) It is imposible, both are measures of variability and they have to provide consistent results.

11. Consider the following six datasets:
A) (1, 5, 6, 10)
B) (4, 4, 4, 4)
C) (101, 102, 103, 104)
D) (7, 8, 9, 10)
E) (1, 2, 9, 10)
F) (1, 2, 3, 4)

a) Which dataset has the smallest SD? B → 1 point
b) Which dataset has the largest SD? E → 1 point
c) Which datasets have the same SD? C, D, and F → 1 point

5. A data set consisting of <u>five</u> numbers has mean, M = 7, and standard deviation, SD = 4. If each of the five numbers is increased by 2, what are the new mean and SD?
E. M = 7, SD = 4
F. M = 9, SD = 4 → 1 point
G. M = 7, SD = 6
H. M = 9, SD = 6

6. An equal number of students competed in the 100m sprint and 100m swim finals. The timings (in seconds) of the champions of the 100m sprint and 100m swim are shown below, as are the average timings and the SDs of the finalists in the two competitions.

	100m sprint	100m swim
Champion	10s	40s
Average of the Finalists, M	12s	45s
SD of the Finalists	1s	3s

Assuming all else being equal, between the two champions, who is the better performer?

E. The sprint champion → 1 point
F. The swim champion
G. Both
H. Not enough information to decide

7. David's scores for Mathematics, Physics and Chemistry in the final examinations are given below. His class's performance for the three subjects is also given below:

	Mathematics	Physics	Chemistry
David's Score	95	90	85
Class Average, M	80	80	80
Class SD	15	5	4

c) Relative to his class, in which subject did David perform the best? Physics → 1 point
d) Relative to his class, in which subject did David perform the worst? Mathematics → 1 point