JoVE JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

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

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 showen 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.
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, $\alpha = .05$, $\beta = .95$,). 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 3, panels A-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, panel 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).

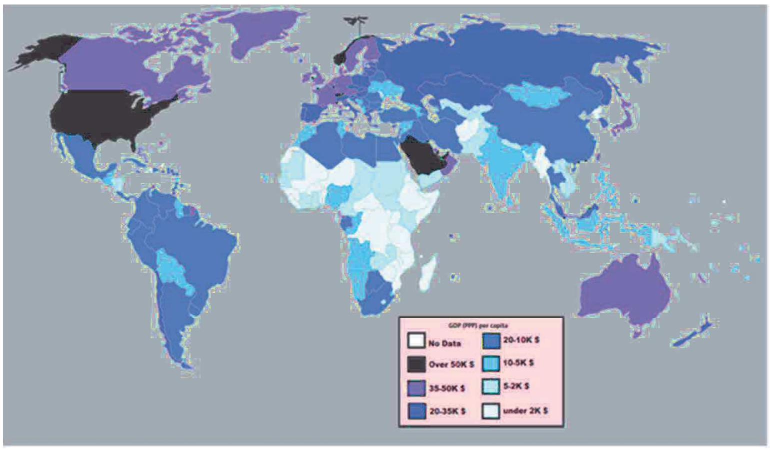
46 Schraw, G. & Dennison, R. S. Assessing metacognitive awareness. *Contemporary educational psychology.* **19** (4), 460-475, (1994).

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).

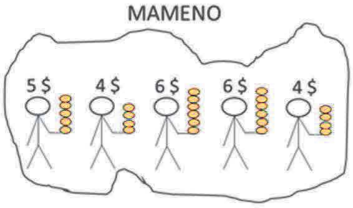


This image shows the average income (GDP per capita) of the countries in 2012. We can see that, for example, Spain and Russia have a similar average income. However, only with the average income, can we have an idea of the wealth of the inhabitants?

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?

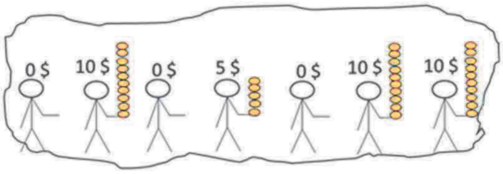
Try to design a mathematical index that would help us to measure and compare inequality in these 4 countries. You can give multiple answers. There is no one unique valid solution!

MAMENO



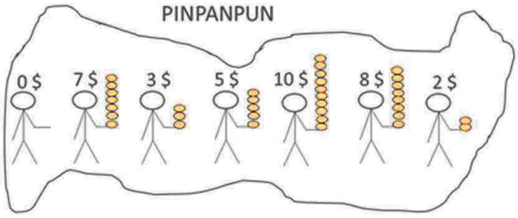
Average Income: (5+ 4 + 6 + 6 + 5) / 5 = **5**

TOVEO



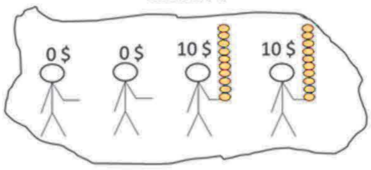
Average Income: (0 + 10 + 0 + 5 + 0 +10 + 10) / 7 = **5**

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.

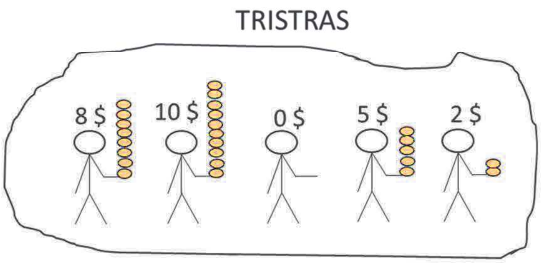


Figure 3: Common Solutions in the Invention Problem and in the Practice Problem.

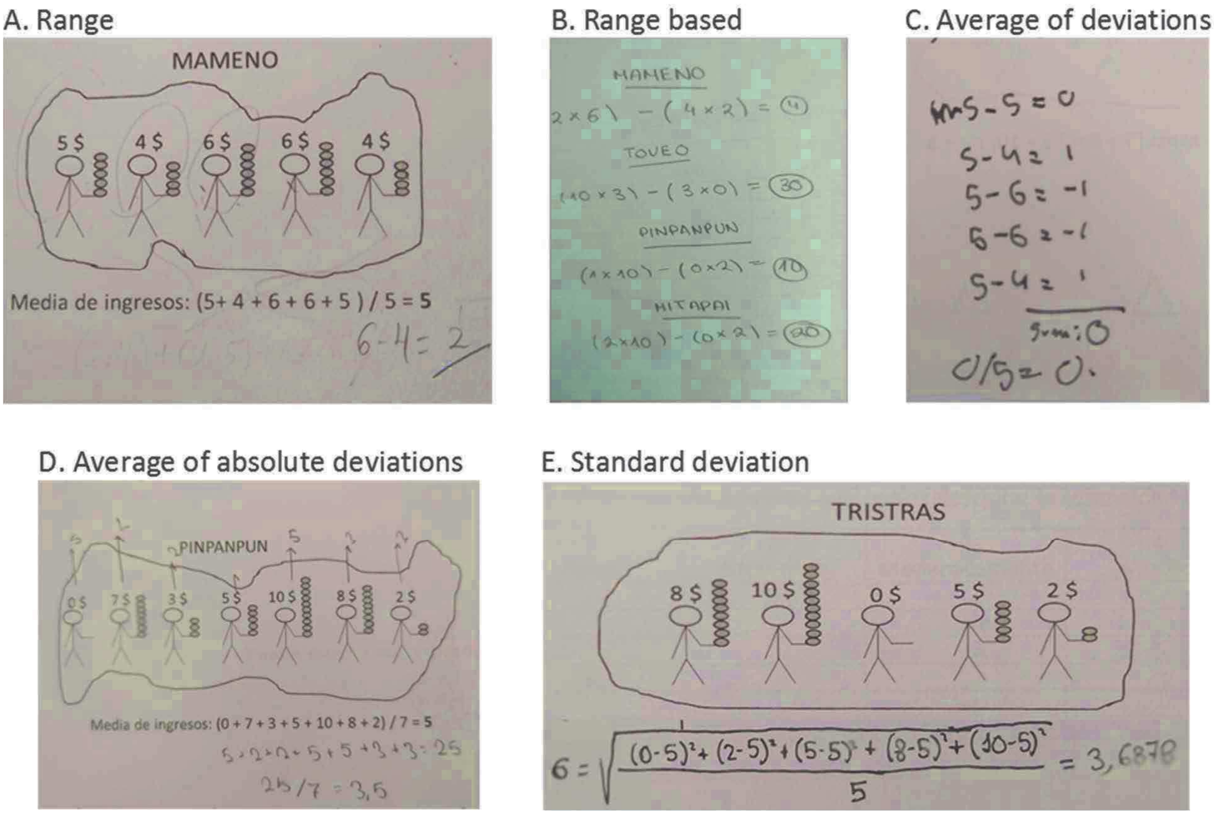


Figure 4: Design of Experimental Evaluation

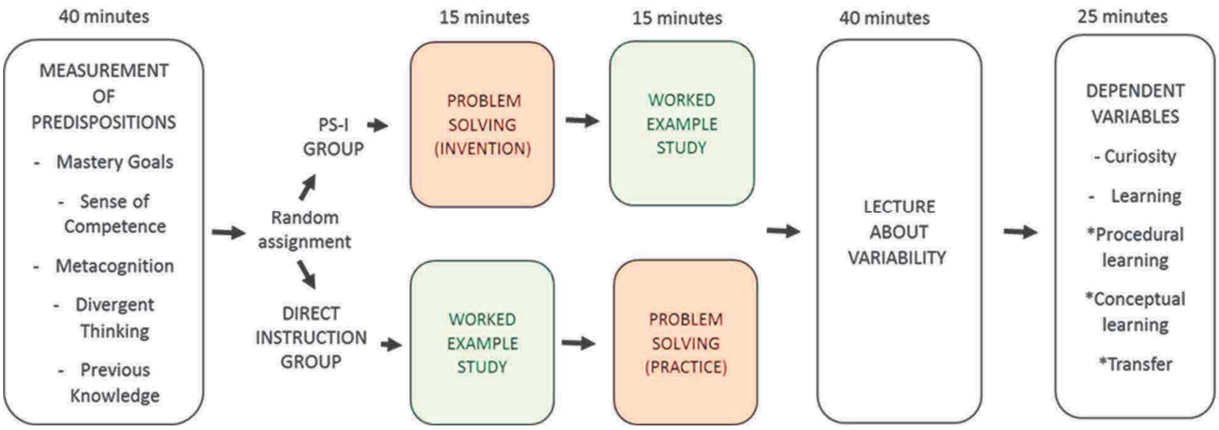
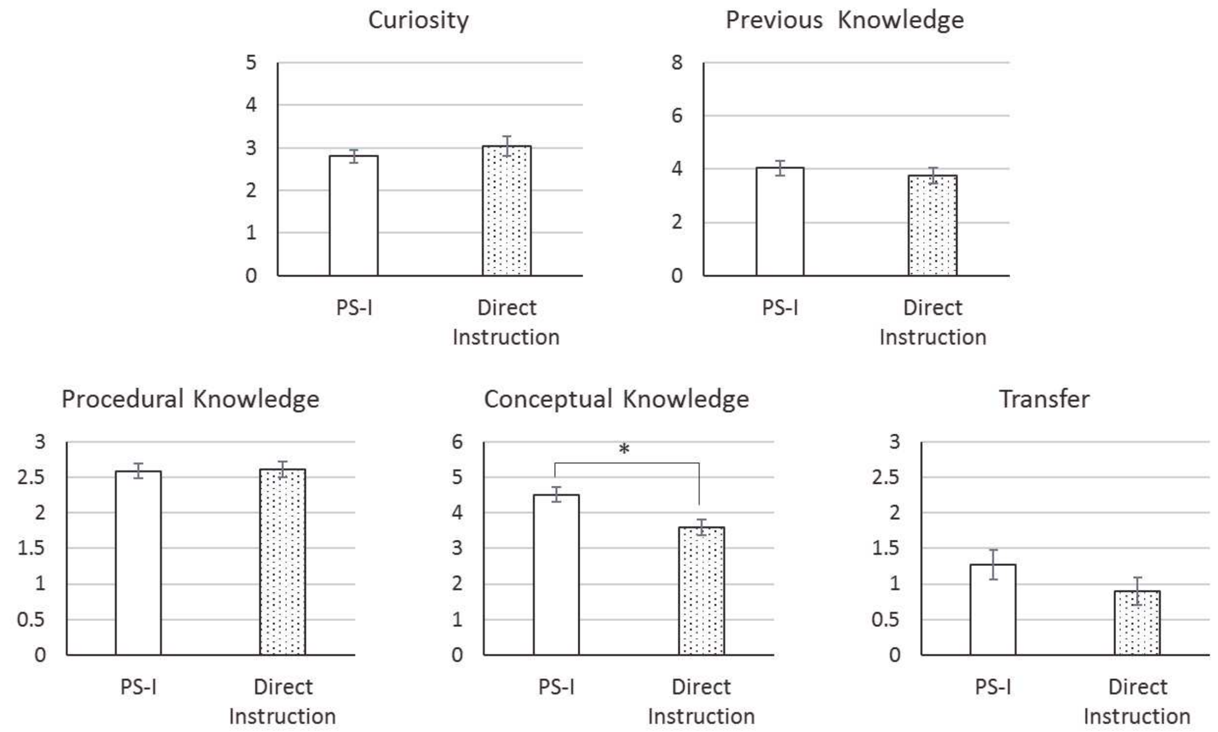
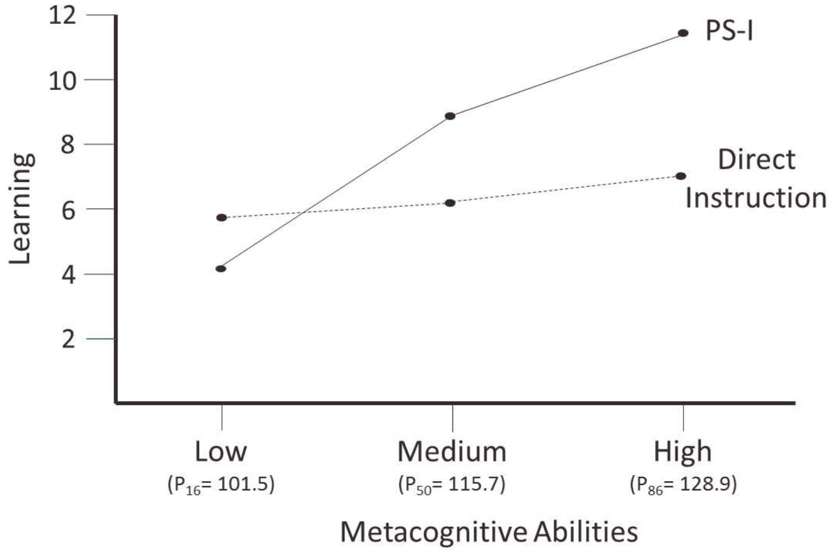


Figure 5: Results about Efficacy of PS-I versus Direct Instruction

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





12
10
8
6
4
2
Learning
PS-I
Direct Instruction
Low
(P16= 101.5)
Medium
(P50= 115.7)
High
(P86= 128.9)
Metacognitive Abilities

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 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

Construct
Curiosity
Learning (procedural, conceptual, and transfer)

Measure and Description
The Curiosity Scale in the Epistemically-Related Emotions Questionnaire 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