Cite the article:

García-Carmona, A., Muñoz, G., Criado, A. M., & Cruz-Guzmán, M. (2024). Validation of an instrument for assessing basic science process skills in initial elementary teacher education. *International Journal of Science Education*, *46*(4), 362-381. https://doi.org/10.1080/09500693.2023.2232936

Validation of an instrument for assessing basic science process skills in initial elementary teacher education

Antonio García-Carmona[1,2,*], Granada Muñoz-Franco[1,3], Ana M. Criado[1,4] and Marta Cruz-Guzmán[1,5]

[1] Departamento de Didáctica de las Ciencias Experimentales y Sociales, Universidad de Sevilla, Spain.

[2] Email: garcia-carmona@us.es; ORCID: https://orcid.org/0000-0001-5952-0340; * corresponding author

[3] Email: gmunoz3@us.es; ORCID: https://orcid.org/0000-0002-6937-1382

[4] Email: acriado@us.es; ORCID: https://orcid.org/0000-0002-6224-9253

[5] Email: mcruzguzman@us.es; ORCID: https://orcid.org/0000-0003-0018-3620

Abstract. Basic science process skills (SPSs) constitute a key dimension of the scientific competence that should be developed from an early age. Therefore, preservice elementary teachers (PETs) who are preparing to teach science should have a sufficient mastery about it. This study presents a questionnaire that allows gradual levels of performance in SPSs to be established, in particular with regard to formulating questions and hypotheses, identifying variables, measurement, planning an inquiry, and interpreting data. The questionnaire was validated in three phases – the first during its design, with the external collaboration of a panel of experts, the second by piloting of a draft versions of the questionnaire on a small sample of PETs, and the third by applying the questionnaire to a sample of 435 PETs. Analysis of the collected data indicated that these were acceptably reliable, and the content and construct validity of the questionnaire was also verified. They also revealed the PETs' generally deficient levels of SPSs. This was especially so regarding identifying suitable questions for school inquiry, understanding precision in measurement, and interpreting data in a table. There was further support for the validity of the questionnaire and the reliability of the outcomes in that it corroborated other results in the literature on the subject that were obtained with different tools and methods.

Keywords: assessment, elementary education, science education, science process skills, teacher education.

Introduction

A task as complex as that of scientifically educating young students (3-12 years) requires teachers of this educational stage to have a series of appropriate teaching competencies (De-Juanas et al., 2016; Howes, 2002). One that stands out is mastery of the school science they have to teach (Kind & Chan, 2019). This includes not only knowledge of the science itself, but also basic *science process skills* (SPSs) (Chabalengula et al., 2012). While it is true that such mastery is usually not the most determining factor in achieving effective science education (Organization for Economic Co-operation and Development [OECD], 2019a), having it is unquestionably an essential requirement for this purpose (Shulman, 1986). Science teachers who do not have a high enough level of preparation in the school science they have to teach will have a limited capacity to create timely and quality learning situations (Garbett, 2003). They are also likely to transmit inadequate conceptions of science to their students (Catalano et al., 2019), and foster poor or misrepresented performance of SPSs.

Education research (García-Carmona & Cruz-Guzmán, 2016; Menon & Sadler, 2016; Verdugo et al., 2016) reveals a low level of scientific competence in trainee teachers of elementary levels of education (pre-service elementary teachers, PETs). There are multiple reasons behind this, but perhaps one of the most notable is that the vast majority of these students access their initial teacher training from academic itineraries with no relation to science (Muñoz et al., 2022). Generally, they had previously had bad experiences with school science (García-Carmona et al., 2014). Therefore, their training in science should begin by assimilating the content of the school science curriculum that has to be taught in elementary education (Cruz-Guzmán et al., 2017; García-Carmona et al., 2017). Only in this way will they later develop the appropriate didactic strategies to effectively implement this school science in their classes (Newman et al., 2004). Indeed, when PETs improve their science content knowledge, their belief in their self-efficacy to be able to teach science also improves (Al Sultan et al., 2018; Menon & Sadler, 2016).

In general, scientific competence comprises the integrated acquisition of skills, attitudes, and scientific and epistemic knowledge (OECD, 2019b). Nonetheless, most studies of PET science education levels have focused on scientific knowledge (e.g., Arslan et al., 2012; Verdugo et al., 2016). Much scarcer have been those aimed at analysing their SPSs (Chabalengula et al., 2012; García-Carmona, 2019, 2020a). For this reason, we set out to design and validate a questionnaire to evaluate the levels of SPSs in PET training.

Theoretical Framework

Science process skills in science education

According to Padilla (1990), SPSs can be defined "as a set of broadly transferable abilities, appropriate to many science disciplines and reflective of the behavior of scientists". Although in science education SPSs have been usually linked to the skills or tasks necessary to carry out a school scientific inquiry (Ash, 1999; Bybee, 2006; Ekici & Erdem, 2020; Harlen, 2014; Wilke & Straits, 2005). It should also be noted that in the literature on science education such abilities/skills are referred to under different names. Thus, for example, while for Padilla (1990) 'measuring' is a science process skill, for Bell (2009) it is a science process/method and for Harlen (2014) it is a science inquiry skill. Similarly, while for Wilke and Straits (2005) 'raising/asking questions' is a scientific method skill, for Harlen (2014) it is a science inquiry skill, and for TIMSS 2019 Science Framework (Mullis & Martin, 2017) and K-12 Framework (National Research Council [NRC], 2012) it is a science practice. The same is true for the other skills. Therefore, although there may be subtle nuances in distinguishing between these different denominations, in this study, we will bring them all together under the umbrella of SPSs.

That said, what is indisputable is that for years there has been a broad international consensus that science education should promote students' acquisition of SPSs as part of their basic scientific literacy (Mullis & Martin, 2017; OECD, 2019b). However, the OECD (1999) emphasized that "processes are only scientific processes when they are used in relation to the subject matter of science." (p. 60). Therefore, SPSs should be addressed in a cross-cutting manner in the context of the different topics of the school science curriculum.

There are a multitude of SPSs proposals in the literature to be addressed in science education, so it would be unaffordable to review all of them here. Yet, some of the most popular or influential ones that would represent the majority of them can be evoked here. For example, more than three decades ago, Padilla (1990) distinguished two types of SPSs: basic and integrated. In the former, he included observing, inferring, measuring, communicating, classifying, and predicting, and in the latter, controlling variables, defining operationally, formulating hypotheses, interpreting data, experimenting, and formulating models. In that same decade, the National Science Education Standards (NRC, 1996) pointed out that, among other SPSs, the activity of inquiry involves making observations, asking questions, planning inquiries, using instruments to acquire, analyse, and interpret data, and communicating the results. Years later, these SPSs were revised and established by NRC (2012) under denomination of science practices (García-Carmona, 2020b). In the same vein, Harlen (2014, p. 13) suggested addressing four groups of SPSs for inquiry-based science education: (a) raising questions, predicting and planning inquiries so as to set up investigations; (b) gathering evidence by observing and using information sources concerned with collecting data; (c) analysing, interpreting, and explaining to be able to draw conclusions; and (d) communicating, arguing, reflecting, and evaluating ideas.

More recently, under the label of science practices, TIMSS 2019 Science Framework (Mullis & Martin, 2017) sets out five SPSs that are fundamental to scientific inquiry: (i) asking questions based on observations, (ii) generating evidence, (iii) working with data, (iv) answering the research question, and (v) making an argument from evidence. Likewise, the PISA theoretical framework for the assessment of scientific competence (OECD, 2019b) and under the denomination of procedural knowledge[1] refers to SPSs such as observation, management and control of variables, measurement, replicability, the abstraction and representation of data, and experimental design in the context of a given question for investigation. PISA justifies the importance of SPSs in epistemological terms by arguing that “Recognising and identifying the features that characterise scientific enquiry requires a knowledge of the standard procedures that underlie the diverse methods and practices used to establish scientific knowledge.” (OECD, 2019b, p. 99). However, in this study we are not concerned with the epistemological dimension of SPSs (i.e., knowledge *about* SPSs), but rather with knowing how to use/apply them in school science learning situations.

In the latest curricular reform for basic science education in Spain, SPSs are also promoted under name of inquiry procedures. Thus, for primary education (6-12 years) it establishes, within a section called Initiation to Scientific Activity, that students must become familiar with (Royal Decree 157/2022):

- 6-8 years: Inquiry procedures appropriate to the needs of the research (observation over time, identification and classification, search for patterns...) (p. 24419).
- 9-10 years: Inquiry procedures appropriate to the needs of the research (observation over time, identification and classification, search for patterns, creation of models, research through information search, experiments with control of variables...) (p. 24423).
- 11-12 years: Phases of scientific inquiry (observation, formulating questions and predictions, planning and conducting experiments, collecting and analysing information and data, communicating results...) (p. 24427).

In addition, for the three cycles of primary education is established the following content (pp. 24419, 24423 and 24427): “Appropriate instruments and devices to make precise observations and measurements according to the needs of the inquiry”.

Similarly, a section called Experimentation in the Environment in the curriculum for early childhood education (0-6 years) includes, as part of basic content, the following references to SPSs (Royal Decree 95/2022): “Control of variables model. Research strategies and techniques: trial-and-error, observation, experimentation, formulation and verification of

[1] However, when PISA was launched, it used the label 'scientific processes' to refer to the same (OECD, 1999).

hypotheses, posing questions, handling and searching in different sources of information." (p. 14584).

Consequently, science education demands, from an early age, explicit attention to SPSs and should therefore occupy a prominent place in PET training plans (Abd Rauf et al., 2013; Rintayati et al., 2020). Because the number of SPSs to be addressed in science education is quite broad, in this study we prioritised those that are most emphasised in the latest curricular prescriptions for basic science education in Spain (Royal Decrees 95/2022 and 157/2022).

Basic science process skills in preservice elementary teacher education

The above sheds light on the SPSs that PETs should acquire in their initial training to teach them in elementary education. In accordance with other authors (Abell et al., 2006; Capps & Crawford, 2013; Newman et al., 2004), the purpose of such training should be for PETs to achieve a sufficient level of mastery of these SPSs (i.e., content knowledge, CK) to then learn to promote them as teachers in elementary students (i.e., pedagogical content knowledge, PCK) (García-Carmona, 2019; García-Carmona et al., 2017). This obviously suggests that the level of comprehension of these SPSs in PETs should be higher than that intended for elementary students, although it is certainly difficult to determine what that level should be. This will depend on several factors, including the familiarity of the PETs with the SPSs, and the characteristics of their educational and training context (Eurydice, 2011). For this reason, we think that a good way to decide this is by consensus of a group of experts on the issue and well aware of the context in which the training plan is to take place, as will be exposed later.

In the literature, numerous studies can be found on PET training in SPSs. Most of them have been conducted in the context of inquiry-based science education (e.g., Abell et al., 2006; Newman et al., 2004; García-Carmona et al., 2017) and reveal that PETs have quite a few limitations in the performance of SPSs. Thus, PETs have difficulties in identifying and controlling variables (Aydoğdu et al., 2014; Criado et al., 2019; Montero-Pau et al., 2017; Rintayati et al., 2020), formulating researchable questions (Cruz-Guzmán et al., 2017), posing scientific hypotheses (García-Carmona, 2020a; Yoon et al., 2012) and planning school inquiries to test them (García-Carmona et al., 2017). PETs also have shortcomings in making measurements (García-Carmona, 2019) and inferring (Aydoğdu et al., 2014; Chabalengula et al., 2012) in the context of a school scientific inquiry. Therefore, it can be said that PETs generally have a rather poor CK about SPSs, and this will inevitably hinder the development of an adequate PCK to teach them in elementary education (Kind & Chan, 2019; Shulman, 1986). So, there is still a long way to go to develop the CK of PETs in relation to SPSs. And from our point of view, a fundamental aspect of this is to make available appropriate instruments to diagnose levels of mastery of basic science skills in PETs.

Assessment of science process skills

In the literature, one finds that students' SPSs are quite frequently evaluated using a qualitative approach, for instance, analysing their productions about an activity of inquiry (Cruz-Guzmán et al., 2017; García-Carmona, 2019, 2020a; Oh, 2010), through class observations and field notes (Yoon et al., 2012), class session recordings (Oh, 2010), or personal interviews (Abd Rauf et al., 2013). This approach has the advantage that it allows the research to go into some depth into the degree of development of these SPSs and the difficulties that hinder their understanding. It is therefore ideal when studying a small group of students. But its main drawback is that it is difficult to handle when one wishes to diagnose the levels of SPSs in a large sample of students.

There are also some questionnaires designed to diagnose SPSs in large samples of students. Examples are the Test of Integrated Process Skills (TIPS II) (Burns et al., 1985), the

Test of Integrated Science Process (TISP) (Shahali & Halim, 2010), and the Test of Basic and Integrated Science Process Skills (T-BISPS) (Al-Junaidi & Ong, 2013). These questionnaires have in common that the items composing them are multiple-choice, so that the subject must choose the correct response among several options. Recently, Šmida et al. (2023) have conducted a comprehensive review of the literature on instruments used to assess SPSs. They also find that the vast majority of them propose a large number of items (some of them include up to 48 items) with a multiple-choice response format. Although these instruments make it possible to easily assess the SPSs of large samples of students, in our opinion, they have several disadvantages. The first is that they allow hardly any depth to be explored regarding students' difficulties on SPSs. Another disadvantage is that there do not make it possible to establish different levels of competency for any one given SPS, given that the evaluation only consists in seeing whether or not the respondent chooses the correct option. What is interesting about an SPSs assessment instrument is that it also allows to determine learning progressions related to them (Pellegrino et al., 2016). Finally, a no-small disadvantage of these instruments is that, due to the large number of items that include, they are difficult to apply in normal science classrooms and therefore difficult to manage for teachers who want to assess the level of acquisition of SPSs in their students. Therefore, it would be interesting to design an instrument that not only diagnoses whether or not students are competent with respect to certain SPSs, but also allows different levels of proficiency to be established in this respect, and that can be easily integrated into the normal development of science teaching.

Aims of the study

Following what has been set out above, we asked ourselves if we could develop a questionnaire that, on the one hand, would be relatively easy to apply to large samples of students (in this case, PETs), and, on the other, would allow differentiation of levels of acquisition for the same SPS. With this in mind, we set the following objectives for the study:

1. Design and validate a questionnaire to diagnose levels of development of SPSs in PETs training.
2. Implement the questionnaire with a large sample of PETs in order to determine their basic SPS levels.

Methods

Participants

To carry out the study, students from the University of Sevilla who were enrolled in the bachelor's degrees in early childhood education (240 credits, deals with stage 0-6 years), primary education (240 credits, deals with stage 6-12 years), and the double degree in early childhood and primary education (364 credits, deals with stage 0-12 years), were invited to participate. They will be referred to here indistinctly as PETs, except when some kind of distinction has to be made as a consequence of the analysis of the results. The three bachelor's degrees include an obligatory subject of science teaching in their training plans. In the primary education and double degrees, the said subject has 9 credits, and in the early childhood education degree, 6 credits. In the three degrees, adapted to the corresponding educational stage, the main purpose of the science teaching subject is for the PETs to: (i) understand the objective of basic science education; (ii) know and analyse the school science curriculum (goals, content, structure, suggestions for teaching, assessment, etc.); (iii) know what difficulties students usually have in learning science; (iv) know resources and methods for science teaching and evaluation; and (v) design activities for teaching/learning science.

The PETs' participation in the study was therefore voluntary and consisted of completing an online questionnaire (Appendix), which included an informative note that the data obtained could be used for a study, under the strict application of ethical standards for research with people. The completion of the questionnaire was proposed just before the PETs began their training in science teaching. The questionnaire was completed by 435 PETs who constituted a convenience sample since those invited were students to whom the researchers had access at the time of the study. Table 1 summarizes the profile of the participants, and represents the set of independent variables of this study.

Table 1. Characteristics of the sample of participants in the study (N = 435)

Distribution by gender:	Women: 80.7% Men: 19.3%
Age:	Range: 18 – 39 years Mean: 19.9 years (standard deviation: 2.3 years)
University degree:	Childhood Education: 11.7% Primary Education: 86.9% Double degree in Childhood and Primary Education: 1.4%
Path to the university degree:	Upper-secondary education in sciences: 15.9% Upper-secondary education in social sciences and humanities: 61.4% Upper-secondary education in arts: 0.5% Vocational training: 20.5% Other (e.g., over 45 years): 1.8%
Self-estimated level of knowledge of science:	Very high: 1.6% High: 14.0% Medium: 53.6% Low: 29.4% Very low: 1.4%

Designing the questionnaire

In the design of the questionnaire, the usual validation and reliability standards for social science research were followed (Drost, 2011). The first author prepared an initial version of the questionnaire with 14 items, in which questions with open responses were combined with mixed questions, i.e., with closed responses to one or several possible options, including a request for justification of the option chosen. In accordance with the literature discussed above and particularly as set out the curricular prescriptions for basic science education in Spain (Royal Decrees 95/2022 and 157/2022), it was decided that the items would cover the following SPSs: identifying research questions through school scientific inquiries, identifying scientific hypotheses, identifying variables in an inquiry, determining classification criteria, planning scientific inquiries, scientific measurements, and data interpretation.

As the first phase of the questionnaire validation process, this initial version of the questionnaire was submitted to scrutiny by a panel of experts. The panel consisted of 12 scholars from eight different Spanish universities who had a proven track record in science education research. They all had extensive experience as instructors of science teaching subjects in initial elementary teacher education. In addition, six of them were or had also been science teachers in secondary education for several years. Basically, the purpose was to apply the three key components of the framework for the validation of an assessment instrument proposed by Pellegrino et al. (2016), namely: *cognitive validity* (what is known about understanding and learning about the content covered to determine the knowledge/skills that students should attain and, therefore, be assessed on), *instructional validity* (how the instrument is aligned with the curriculum and other didactic standards in relation to the content being addressed), and *inferential validity* (how the instrument provides reliable, accurate, and analysable information to diagnose cognitive states of learners with respect to the content addressed).

Consequently, for the evaluation of the questionnaire, the experts had to (1) examine the content of the items according to the objective pursued with them, (2) propose an expert response for each item reflecting the desirable level of mastery for PETs about the SPS considered, (3) review the wording of the items and formulate, where appropriate, modifications to improve their comprehensibility, and (4) suggest any question they deemed appropriate as an alternative or addition to those of the proposal. As a result of this process, it was decided to remove from the questionnaire those items that were not agreed upon by at least three-quarters of the panel (i.e., 9 out of the 12 experts) regarding the suitability of their content, their wording, and the expert responses that were proposed for them. Thus, a second 7-item version of the questionnaire was obtained, with wording that included the improvements suggested by the experts. Furthermore, in this version of the questionnaire, the SPS related to the determination of classification criteria was no longer present because, due to the open nature of the item, multiple possible responses emerged, making it difficult to determine an unambiguous rubric for their classification. Concerning the response considered most appropriate for each item, basically the one that would be required of a student finishing compulsory education (i.e., 16 years) was assumed. It was a reasonable decision bearing in mind that, in Spain, many of the students entering elementary teacher education last studied science at 14-15 years of age. For the same reason, the contexts used for the items were also intended to be at least recognisable to a secondary school student.

The next phase validation consisted of applying the second version of the questionnaire, as a pilot test, to a convenience sample of 37 primary teacher education students, chosen for being available during this phase of the study. The questionnaire was entered into the Google Form application with the purpose of creating a link to enable the students to complete the questionnaire online, with their responses being automatically entered into a spreadsheet. Once the responses had been obtained, those of 12 randomly selected students (32.3%) were categorized separately by each of the four researchers, following the usual guidelines in qualitative content analysis (Mayring, 2000). To this end, the researchers applied a level categorization rubric that had been designed a priori.

In pooling the results of the four researchers' categorizations, majority (3 out of 4) or full (4 out of 4) agreement was found in 82.1% of the cases. Likewise, the data were subjected to an inter-rater agreement analysis by calculating the Fleiss' kappa index. For a 95% asymptotic significance level, this index was found to be 0.60, which can be considered as between moderate and good (Laerd Statistics, 2019). This pilot application of the second version of the questionnaire made it possible to improve the wording and typology of the items, as well as the rubric for categorizing the responses. For example, included as distractors in some items were inappropriate or incomplete responses which had been observed as trends in the pilot application. The version resulting from all this analysis was finally passed on to the large PETs sample described in the previous subsection.

Data collection and analysis

The final version of the questionnaire was completed by the participants online, through a link created for this purpose. The responses were recorded on an Excel spreadsheet. Subsequently, they were coded and exported to the SPSS v.26 program in order to perform the corresponding statistical analyses.

The data were obtained from an evaluation rubric with which the responses were categorized at different levels of proximity to the ideal response expected for each item (Table 2). As explained above, its design followed a dynamic process (Cáceres, 2003) in which two stages should be highlighted. The first took place throughout the process of designing and validating the questionnaire, with the categories being refined until obtaining a rubric that would be applicable to the data obtained in the final application (a priori design). The second was carried out once the data had been obtained (a posteriori design), with some final adjustments being made as a result of discussions between the researchers to definitively categorize the responses.

Although the questionnaire is made up of some items with options to choose from, the rubric allows the SPSs to be classified into four levels, with 4 representing the maximum desirable level, and 1 an inadequate or very deficient level. The other levels denote intermediate attainment in the development of SPSs. Thus, for example, in Item 6, option (e) represents a Level 3 because, even though what it indicates is right, its description is not as complete or well-fitted to the question as that expressed in option (b) (Level 4, maximum).

Table 2. Rubric for the analysis of the items in the questionnaire

Science process skills	Items	Rubric
Identifying questions that are researchable through school inquiries	1	Level 4 (maximum): Selects only the FOUR questions that would be researchable in the context of a school science inquiry: (a), (c), (e), and (h). Level 3: Selects THREE of the four Level 4 questions, and no non-researchable questions: (b), (d), (f), (g) and (i). Level 2: Selects THREE or FOUR questions from Level 4, and one non- researchable question. Level 1 (minimum): Selects TWO or more non-researchable questions, regardless of the number of researchable questions got right.
Identifying scientific hypotheses	2	Level 4 (maximum): Selects only options (b) and (d) as scientific hypotheses. Level 3: Selects ONLY option (b) or ONLY option (d) as scientific hypothesis. Level 2: Select options (b) and (d), but also option (a) or (c) as scientific hypotheses. Level 1 (minimum): Other possibilities.

Identifying variables in a scientific inquiry	3	Level 4 (maximum): Identifies the independent (IV) and dependent (DV) variables adequately in the four questions: (a) IV: temperature; DV: pressure (b) IV: medium (cotton-wool/soil); DV: growth/height (* "time to reach a certain height", "growth reached in a certain period of time", or "speed/rapidity of growth" are considered valid) (c) IV: thickness; DV: resistance/intensity (d) IV: mass/weight; DV: time Level 3: Identifies the IV and DV adequately in three of the four cases. Level 2: Identifies the IV and DV adequately in one or two of the four cases. Level 1 (minimum): Does not adequately identify the IV and DV in any of the cases.
Planning a scientific inquiry	4	Level 4 (maximum): option (d) Level 3: option (a) Level 2: option (e) Level 1 (minimum): options (b) or (c)
Selecting measurement instruments	5	Level 4 (maximum): option (a) Level 3: option (b) Level 2: option (d) Level 1 (minimum): options (c) or (e)
Interpretation of a data graph	6	Level 4 (maximum): option (b) Level 3: option (e) Level 2: option (d) Level 1 (minimum): options (a) or (c)
Interpretation of a data table	7	Level 4 (maximum): option (c) Level 3: option (a) Level 2: option (e) Level 1 (minimum): options (b), (d) or (f)

In addition to the descriptive analysis of SPSs, the possible influences were explored of the participants' profiles (independent variables) on the responses (dependent variables). To this end, the data were first checked for statistical normality using the Kolmogorov-Smirnov test. The result indicated that, for a significance level of 0.05, the data were not normally distributed. Therefore, the correlation analyses of the variables had to be carried out with non-parametric statistical tests. In particular, the chi-squared test (χ^2) was applied to determine the significance of any possible correlations. In order to expedite the presentation of results, only those correlations that were statistically significant will be indicated.

Results

Reliability of data and validity of the questionnaire

The questionnaire developed has a series of peculiarities that distinguish it from other more traditional questionnaires that use ordinal scales. In this case, the gradual levels of responses to some items arise from a combination of the results observed in the different parts of the same item (see the rubric for Items 1, 2, and 3, Table 2). The analysis of the reliability of the data was thus more complex than for those other types of questionnaires.

In order to determine the reliability of the categorization process of the responses, an analysis of inter-rater agreement and concordance was made (Gisev et al., 2013; Lange, 2011). For this, the second author carried out a complete coding of the responses in accordance with the category levels of the a priori rubric. This coding was later reviewed by the first author. The process resulted in full agreement for 97.5% of the cases, and a value of Cohen's kappa equal to 0.82, which indicates quite good agreement (Abraira, 2001). The cases of discrepancy were discussed together with the other two researchers until consensus agreement was reached. These agreements also had an impact on the final design of the rubric, as indicated above.

On the other hand, the fact that the questionnaire consists of a small number of items with different structures and several distinguishable dimensions a priori (i.e., different types of SPSs) suggested that estimating the internal consistency of results through statistics such as Cronbach's alpha was not an appropriate strategy (Eisinga et al., 2013; Oviedo & Campo-Arias, 2005; Pallant, 2011). Even so, it was calculated, and a very low alpha value was indeed obtained,[2] which confirmed that the items designed measure different constructs (in this case different SPSs) and therefore form a multidimensional questionnaire. In such a situation, what advisable is to make a factor analysis of the items to determine how they would be grouped into different latent dimensions (Frías-Navarro, 2022). This factor analysis was made, and it was also useful in determining the construct validity of the questionnaire, as will be seen below.

With regard to the questionnaire's content validity, it can be said that this was quite guaranteed as a result of the process followed in its design with the help of the panel of experts (Escobar-Pérez & Cuervo-Martínez, 2008). Furthermore, to determine its construct validity, the data were subjected to a factor analysis (Lloret-Segura et al., 2014). This included different SPSs which could be grouped into different dimensions. However, these groupings were not established a priori; thus, an exploratory factor analysis of principal component was carried out followed by an oblique rotation (Oblimin) with Kaiser normalization, as is suggested for social science studies (Frías-Navarro & Pascual, 2012).

The factor analysis gave a KMO measure of sampling adequacy equal to 0.54 (greater than 0.5, which is the value above which it is considered acceptable) and a Bartlett's sphericity value of 0.006 (< 0.05), which indicated that the test was statistically relevant (IBM, 2022). In addition, the test yielded a factorial model that clustered the items into three components, explaining 49.3% of the total variance (14.6%, 15.6% and 19.1% respectively). These factorial components are presented in Table 3, labelled with a brief overall description of their factors. As can be observed, all items have a factor loading in absolute value equal to or greater than 0.5. According to Hair et al. (2010), this indicates that the items have an acceptable correlation with the factor or dimension in which they are included. Moreover, such a grouping of items is acceptable theoretically, i.e., to put it in another way, the grouping does not imply any kind of contradiction or conflict within the framework of SPSs (although in theory other groupings could be made as well). Therefore, its assumption provides the questionnaire with acceptable construct validity, and this grouping will underpin how the results will be presented later.

---

[2] According to Oviedo and Campo-Arias (2005), ideally Cronbach's alpha should only be calculated for unidimensional scales with between three and twenty items.

It should be also noted that items 1 and 5 in component 1 (‘Initiation of an inquiry and selection of instruments’) have negative factor loadings. This indicates that these items are inversely correlated with this component factor or dimension. One possible explanation for this is that the responses with the lowest levels of adequacy in these items, according to the rubric, are the ones that determine that they can be grouped into this dimension with a considerable correlation.

Table 3. Principal components resulting from the exploratory factor analysis with oblique rotation (Oblimin)* to determine the questionnaire’s construct validity

	1. Initiation of an inquiry and selection of instruments	2. Inquiry methods	3. Data interpretation
Identifying researchable questions (Item 1)	-0.62		
Identifying scientific hypotheses (Item 2)	0.62		
Selecting measurement instruments (Item 5)	-0.5		
Planning a scientific inquiry (Item 4)		0.74	
Identifying variables (Item 3)		0.64	
Interpretation of a data graph (Item 6)			0.8
Interpretation of a data table (Item 7)			0.58

* *Rotation has converged in 14 iterations*.

Science process skills of trainee elementary education teachers

Initiation of an inquiry and selection of instruments

The results of Item 1 are presented in Table 4. It can be seen that slightly more than two-thirds of the PETs fail to differentiate questions that are researchable through school scientific inquiries from those that are not (Level 1). Only a proportion slightly greater than 10% of the sample distinguished them clearly (Level 4). Also, small percentages of the PETs (together, just over 20%) are at the intermediate response levels (Levels 2 and 3), denoting doubts in their identification of and/or distinction between researchable and non-researchable questions.

Table 4. Identification of researchable questions (Item 1)

Response level		Brief description of the level
Level 4:	10.6%	Clearly identifies and distinguishes all questions that are researchable through school scientific inquiries from those that are not.

Level 3:	5.5%	Identifies a majority of the researchable questions and rules out those that are not.
Level 2:	14.7%	Identifies a majority of the researchable questions, but does not rule out some that are not.
Level 1:	69.2%	Does not distinguish between researchable and non-researchable questions.

Slightly better results were obtained on Item 2 (Table 5). About a third of the PETs clearly identified what a scientific hypothesis is (Level 4). Nonetheless, almost 41% of the PETs were at Level 1, which indicates that they did not recognize the characteristic features of a hypothesis in the context of a scientific inquiry. Intermediate response levels, which reflect a somewhat limited (Level 3) or confused (Level 2) understanding of hypotheses, corresponded to around a quarter of the PETs.

Table 5. Identification of scientific hypotheses (Item 2)

Response level		Brief description of the level
Level 4:	35.4%	Clearly identifies the two scientific hypotheses, in which a tentative explanation f the researchable question posed is given, the veracity of which can be checked.
Level 3:	17.0%	Only identifies one of the two scientific hypotheses.
Level 2:	6.9%	Despite identifying the two scientific hypotheses, does not rule out any of the one which are not.
Level 1:	40.7%	Does not distinguish a scientific hypothesis from other types of scientific stateme

With respect to scientific measurement (Table 6), about 27% of the sample selected the instrument that provided the greatest possible precision (Level 4), while 44.6% of the PETs did not take this concept into account when selecting the most suitable instrument to carry out the measurement (Level 1). The intermediate response levels (Levels 2 and 3) once again covered more than a quarter of the PETs.

Table 6. Selection of measurement instruments (Item 5)

Response level		Brief description of the level
Level 4:	27.6%	Selects the instrument that will allow the measurement to be made with the greatest possible accuracy.
Level 3:	10.1%	Selects the second most accurate measuring instrument, assuming that none of those proposed allows the measurement to be made in one go.
Level 2:	17.7%	Selects the instrument that will allow the measurement to be made in one go, although it is not the most accurate.
Level 1:	44.6%	Selects the instruments with the lowest accuracy, without taking into account the importance of this concept in scientific measurement.

Inquiry methods

The identification of variables in scientific inquiries (Table 7) is also a major difficulty for the PETs. Around 94% of the sample were included in the lowest response levels (Levels 1 and 2). The percentage of PETs capable of identifying the independent and dependent variables in all the inquiries described (Level 4) did not reach 1%, and only a little more than 5% managed to do so in most cases.

The correlation analysis of variables also indicated that, for $p < 0.01$, there were significant differences in the results of this skill with respect to the academic route preceding access to the degree course ($\chi^2 = 102.09$, df = 12, $p = 0.000$) and the level of self-estimated science knowledge ($\chi^2 = 30.64$, df = 12, $p = 0.002$). With respect to the former, the best results were among those PETs who came from science upper-secondary education. And regarding the latter, this correlation was favourable to those PETs with higher levels of self-estimated knowledge of science.

Table 7. Identification of variables (Item 3)

Response level		Brief description of the level
Level 4:	0.7%	Identifies and clearly distinguishes the independent (IV) and dependent (DV) variables in all the inquiries described.
Level 3:	5.3%	Clearly identifies and distinguishes IV and DV in most of the inquiries described.
Level 2:	57.5%	Identifies and distinguishes IV and DV only in a minority of the inquiries described.
Level 1:	36.6%	Does not adequately identify or distinguish IV and DV in any of the inquiries described.

When it comes to identifying which inquiry, plan is the most appropriate for a problem (Table 8), the positive results are better than in identifying the variables. It was found that around 45% of the PETs chose the most appropriate option (Level 4), and around 20% chose the one that is closest (Level 3). Even so, around a third of the sample showed significant deficiencies regarding this SPS.

Table 8. Planning a scientific inquiry (Item 4)

Response level		Brief description of the level
Level 4:	45.3%	Identifies the research problem and a specific and reliable method to address it.
Level 3:	20.7%	Identifies the inquiry problem, but not its necessary relationship with a particular method to address it.
Level 2:	26.0%	Only identifies some elements of a possible method for inquiry, which are unrelated to the problem at hand.
Level 1:	8.0%	Does not identify the problem or any method to address it.

Data interpretation

Regarding the interpretation of data from an information source, if we look at the highest response levels, we can see that the PETs seem to deal somewhat better with a data table (Table 9) than with a graph (Table 10). Indeed, while slightly more than a third of the PETs managed to justify their response to the problem with a reasonably adequate handling of the data (Levels 3 and 4), in the case of the data table, the proportion of higher levels was close to half the sample. Nonetheless, it should be noted that, in both situations in which data interpretation is dealt with, the lowest response levels (Levels 1 and 2) accounted for a high percentage of the PETs – more than 65% in the case of the graph, and close to 51% in that of the data table.

Table 9. Interpretation of a data graph (Item 6)

Response level		Brief description of the level
Level 4:	13.6%	Adequately interprets the data in the graph and uses it as an argument to respond in the most accurate way possible to the question posed.
Level 3:	20.9%	Adequately interprets the data in the graph, but the response does not fully fit what is being asked.
Level 2:	48.1%	Answers the question qualitatively, but neither referring to specific data in the graph nor giving any explanation.
Level 1:	17.5%	Interprets graph data inappropriately.

On the other hand, significant differences, at $p < 0.05$, were found in the results of Item 6 according to the PETs' levels of self-estimated science knowledge ($\chi^2 = 21.08$, df = 12, $p = 0.049$). Specifically, the students with higher self-estimated science levels performed better in interpreting the data in the graph. Similarly, differences were found in Item 7, $p < 0.05$, in relation to the degree the PETs were studying for ($\chi^2 = 9.78$, df = 4, $p = 0.044$). The PETs who were doing a primary education degree showed better skills in interpreting a data table than those of the other degree courses (degree in childhood education and double degree). A significant correlation was also found, $p < 0.01$, regarding the route the PETs took to access the degree ($\chi^2 = 23.36$, df = 8, $p = 0.000$). The higher levels in this SPS were achieved by the PETs who came from science upper-secondary education.

Table 10. Interpretation of a data table (Item 7)

Response level		Brief description of the level
Level 4:	34.3%	Properly interprets the data in the table and uses it as an argument to answer the question posed.
Level 3:	15.2%	Identifies the data in the table well, but makes a somewhat limited interpretation of it when answering the question posed.
Level 2:	0.0%	Only makes a qualitative interpretation of the data in the table to answer the question posed.
Level 1:	50.6%	Interprets the data in the table inappropriately.

Discussion

This study provides an alternative instrument for evaluating SPSs in PETs training. The meticulous procedure followed in designing the questionnaire and the associated rubric for categorizing the responses, as well as the statistical analyses carried out of the resulting data, reveal that the instrument has sufficient validity to diagnose SPSs in a reliable way. This is an interesting result, since it is a novel questionnaire in comparison with previous ones (e.g., Al-Junaidi & Ong, 2013; Burns et al., 1985; Shahali & Halim, 2010), which are usually limited to the use of a multiple choice scale, with nominal variables, where the respondent must choose the correct response among several possible options. The questionnaire presented here, however, allows one to determine different levels of performance for each of the SPSs asked. This not only provides a more refined diagnosis but can also be very useful in more accurately designing specific activities to improve the SPS levels observed in the PETs.

In relation to the PETs' SPSs, this study confirms what the literature has repeatedly advanced – that, in general, PETs undertake their training in science teaching with rather limited scientific competence (Garbett, 2003; García-Carmona & Cruz-Guzmán, 2016; Newman et al., 2004). In this case, especially noteworthy difficulties were found regarding the identification of questions researchable through school inquiries, the accuracy of a measurement, and the interpretation of a data table. Therefore, the PETs start from a situation that is rather weak for acquiring adequate pedagogical content knowledge to be able to teach science (Garbett, 2003; Kind & Chan, 2019; Shulman, 1996).

Although the above is an unsatisfactory outcome, it also allows for a positive reading. Indeed, the fact of confirming what was expected, unfortunately, in light of the vast amount of research carried out with PETs, indicates that the questionnaire presented has acceptable validity and that the data attained are reliable. Specifically, the questionnaire allowed corroboration of results obtained in previous qualitative studies on the performance of PETs in relation to SPSs, such as identifying researchable questions (Cruz-Guzmán et al., 2017), formulating scientific hypotheses (García-Carmona, 2020a; Yoon et al., 2012), taking measurement accuracy into account (García-Carmona, 2019), planning an inquiry (García-Carmona et al., 2017), identifying and handling variables (Aydoğdu et al., 2014; Criado et al., 2019; Montero-Pau et al., 2017), and handling and interpreting data (Aydoğdu et al., 2014; García-Carmona, 2020a).

Another important aspect that can be derived from this study is that those students who access their initial teacher training from science upper-secondary education show a significantly better performance in some SPSs than their peers from other pre-university degrees unrelated to science. This suggests that science upper-secondary education should perhaps be established as a priority academic itinerary to access a university degree aimed at training PETs.

On the other hand, it is necessary to refer to the limitations of this study. The first is related to the sample of participants. Having been chosen for convenience, the results are not generalizable to the entire PET population. Nonetheless, the sample was quite large and, according to the profiles of the participants, quite similar to the usual PET population in Spain (Muñoz et al., 2022). Therefore, the results may to a large extent be extrapolable to PETs studying at other Spanish universities, as well as to those from other countries with similar characteristics.

The second limitation that should be highlighted is the not very large number of SPSs that were analysed. This was the result of the rigorous validation process followed in the design of the questionnaire, which sought to select only those items that had the greatest consensus from the panel of experts consulted. In this sense, the study was undertaken with strong guarantees of the questionnaire's validity. But this suggests the need to expand the questionnaire by adding further SPSs, with items that are valid and reliable, so as to make useful diagnoses of their performance levels.

A third limitation that could be highlighted is the possibility that the phenomena and contents of school science, chosen as contexts in the design of the items, would have implied an added difficulty for the PETs' responses. In other words, one might ask whether the same results would have been obtained if other phenomena or school science content had been used as contexts to evaluate the same skills. Although it is inevitable to use a specific phenomenological context to ask about a particular SPS (OECD, 1999), it would be interesting to design and validate new items on the same SPSs with contexts different from those used in the current questionnaire. This would also help strengthen the reliability of the data obtained with the questionnaire in terms of its internal consistency. Since the calculation of coefficients such as Cronbach's alpha did not seem appropriate in this case, as discussed above, the inclusion in the questionnaire of different items on the same SPSs would allow us to compare their responses by means of the analysis known as the split-half method.

Finally, it is worth asking how the designed questionnaire might work with science learners of more basic educational levels, such as those in secondary education. After all, in the design and validation of the questionnaire, the level and demands of school science in the latter stage were very much present in the design and validation of the questionnaire. This and the implications of the aforementioned limitations constitute challenges that should be addressed in future research on the subject.

Funding

This study was funded by MCIN/AEI/10.13039/501100011033 (Government of Spain) under Grant EDU2017-82505-P.

Ethical Statement

This study met the ethics requirements for research that involves human subjects at the time the data was collected.

References

Abd Rauf, R. A., Rasul, M. S., Mans, A. N., Othman, Z., & Lynd, N. (2013). Inculcation of science process skills in a science classroom. *Asian Social Science*, *9*(8), 1911-2017. http://dx.doi.org/10.5539/ass.v9n8p47

Abell, S. K., Smith, D. C., & Volkmann, M. J. (2006). Inquiry in science teacher education. In L. B. Flick & N. G. Lederman (Eds.), *Scientific inquiry and nature of science* (pp. 173-199). Dordrecht: Springer.

Abraira, V. (2001). El índice kappa [Kappa index]. *Semergen*, *27*(5), 247-249. https://doi.org/10.1016/S1138-3593(01)73955-X

Al Sultan, A., Henson, H., & Fadde, P. J. (2018). Pre-service elementary teachers' scientific literacy and self-efficacy in teaching science. *IAFOR Journal of Education, 6*(1). https://doi.org/10.22492/ije.6.1.02

Al-Junaidi, M., & Ong, E. T. (2013). Test of basic and integrated science process skills (T-BISPS): How do form four students in Kelantan Fare? *Asian Journal of Assessment in Teaching and Learning*, *3*, 15-30. https://ejournal.upsi.edu.my/index.php/AJATeL/article/view/1943

Arslan, H. O., Cigdemoglu, C., & Moseley, C. (2012). A three-tier diagnostic test to assess pre-service teachers' misconceptions about global warming, greenhouse effect, ozone layer depletion, and acid rain. *International Journal of Science Education*, *34*(11), 1667-1686. https://doi.org/10.1080/09500693.2012.680618

Ash, D. (1999). The Process Skills of Inquiry. In *Inquiry: Thoughts, Views, and Strategies for the K-5* (pp. 51-62). National Science Foundation.

Aydoğdu, B., Erkol, M., & Erten, N. (2014). The investigation of science process skills of elementary school teachers in terms of some variables: Perspectives from Turkey. *Asia-Pacific Forum on Science Learning & Teaching, 15*(1), 1-28. https://www.eduhk.hk/apfslt/download/v15_issue1_files/baydogdu.pdf

Bell, R. L. (2009). Teaching the nature of science: Three critical questions. In *Best Practices in Science Education*. National Geographic School Publishing.

Burns, J. C., Okey, J. R., & Wise, K. C. (1985). Development of an integrated process skill test: TIPS II. *Journal of Research in Science Teaching, 22*(2), 169–177. https://doi.org/10.1002/tea.3660220208

Bybee, R. W. (2006). Scientific inquiry and science teaching. In L. B. Flick & N. G. Lederman (Eds.), *Scientific inquiry and nature of science* (pp. 1–14). Springer.

Cáceres, P. (2003). Análisis cualitativo de contenido: Una alternativa metodológica alcanzable [Qualitative content analysis: an achievable methodological alternative]. *Psicoperspectivas, Individuo y Sociedad*, *2*(1), 53–82. https://dx.doi.org/10.5027/psicoperspectivas-Vol2-Issue1-fulltext-3

Capps, D. K., & Crawford, B. A. (2013). Inquiry-based professional development: What does it take to support teachers in learning about inquiry and nature of science? *International Journal of Science Education, 35*(12), 1947–1978. https://doi.org/10.1080/09500693.2012.760209

Catalano, A., Asselta, L., & Durkin, A. (2019). Exploring the relationship between science content knowledge and science teaching self-efficacy among elementary teachers. *IAFOR Journal of Education*, *7*(1), 57-70. https://files.eric.ed.gov/fulltext/EJ1217961.pdf

Chabalengula, V. M., Mumba, F., & Mbewe, S. (2012). How pre-service teachers understand and perform science process skills. *Eurasia Journal of Mathematics, Science and Technology Education*, *8*(3), 167-176. https://doi.org/10.12973/eurasia.2012.832a

Criado, A. M., Illescas-Navarro, M., & Muñoz, G. (2019). Diagnosis of difficulties in of handing variables and drawing conclusions by means of the reports of an initial experimental activity. In *Proceedings of Educating the Educator III* (pp. 101-102). Freiburg, Germany. https://icse.eu/wp-content/uploads/2019/10/Conference-Book-ETE-III_Final.pdf

Cruz-Guzmán, M., García-Carmona, A., & Criado, A. M. (2017). An analysis of the questions proposed by elementary pre-service teachers when designing experimental activities as inquiry. *International Journal of Science Education*, *39*(13), 1755-1774. https://doi.org/10.1080/09500693.2017.1351649

De-Juanas, A., Martín, R., & Pesquero, E. (2016). Teaching competences necessary for developing key competences of primary education students in Spain: teacher assessments. *Teacher Development*, *20*(1), 123-145. https://doi.org/10.1080/13664530.2015.1101390

Drost, E. A. (2011). Validity and reliability in social science research. *Education Research and Perspectives, 38*(1), 105-123. https://search.informit.org/doi/10.3316/ielapa.491551710186460

Eisinga, R., Grotenhuis, M. T., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, *58*, 637-642. http://dx.doi.org/10.1007/s00038-012-0416-3

Ekici, M., & Erdem, M. (2020). Developing science process skills through mobile scientific inquiry. *Thinking Skills and Creativity, 36*, 100658. https://doi.org/10.1016/j.tsc.2020.100658

Escobar-Pérez, J., & Cuervo-Martínez, A. (2008). Validez de contenido y juicio de expertos: una aproximación a su utilización [Content validity and expert judgement: an approach to their use]. *Avances en Medición*, *6*(1), 27-36. http://www.humanas.unal.edu.co/psicometria/files/7113/8574/5708/Articulo3_Juicio_de_expertos_27-36.pdf

Eurydice. (2011). *Science education in Europe: National policies, practices, and research*. Publications Office of the European Union.

Frías-Navarro, D. (2022). Apuntes de estimación de la fiabilidad de consistencia interna de los ítems de un instrumento de medida [Notes on estimating the internal consistency reliability of the items of a measuring instrument]. Universidad de Valencia. https://www.uv.es/friasnav/AlfaCronbach.pdf

Frías-Navarro, D., & Pascual, M. (2012). Prácticas del análisis factorial exploratorio (AFE) en la investigación sobre conducta del consumidor y marketing [Exploratory factor analysis (EFA) in consumer behavior and marketing research]. *Suma Psicológica*, *19*(1), 47-58. https://www.redalyc.org/articulo.oa?id=134224283004

Garbett, D. (2003). Science education in early childhood teacher education: Putting forward a case to enhance student teachers' confidence and competence. *Research in Science Education*, *33*(4), 467-481. https://doi.org/10.1023/B:RISE.0000005251.20085.62

García-Carmona, A. (2019). Pre-service primary science teachers' abilities for solving a measurement problem through inquiry. *International Journal of Science and Mathematics Education*, *17*(1), 1-21. https://doi.org/10.1007/s10763-017-9858-7

García-Carmona, A. (2020a). Prospective elementary teachers' abilities in tackling a contextualized physics problem as guided inquiry. *Revista Brasileira de Ensino de Física, 42*, e20190280. https://doi.org/10.1590/1806-9126-RBEF-2019-0280

García-Carmona, A. (2020b). From inquiry-based science education to the approach based on scientific practices. *Science & Education*, *29*(2), 443-463. https://doi.org/10.1007/s11191-020-00108-8

García-Carmona, A., & Cruz-Guzmán, M. (2016). ¿Con qué vivencias, potencialidades y predisposiciones inician los futuros docentes de Educación Primaria su formación en la enseñanza de la ciencia? [What personal experiences, potentialities and predispositions do prospective primary teachers manifest when they start their training in science teaching?] *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias, 13*(2), 440-458. http://hdl.handle.net/10498/18299

García-Carmona, A., Criado, A. M., & Cruz-Guzmán, M. (2017). Primary pre-service teachers' skills in planning a guided scientific inquiry. *Research in Science Education, 47*(5), 989–1010. https://doi.org/10.1007/s11165-016-9536-8

García-Carmona, A., Cruz-Guzmán, M., & Criado, A. M. (2014). "¿Qué hacías para aprobar los exámenes de ciencias, qué aprendiste y qué cambiarías?". Preguntamos a futuros docentes de Educación Primaria ["What did you do to pass the science exams, what did you learn and what would you change?" We ask future teachers of Primary Education]. *Investigación en la Escuela, 84*, 31–46. https://doi.org/10.12795/IE.2014.i84.03

Gisev, N., Bell, J. S., & Chen, T. F. (2013). Interrater agreement and interrater reliability: key concepts, approaches, and applications. *Research in Social and Administrative Pharmacy*, *9*(3), 330-338. https://doi.org/10.1016/j.sapharm.2012.04.004

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Pearson.

Harlen, W. (2014). Helping children's development of inquiry skills. *Inquiry in Primary Science Education*, *1*(1), 5-19. https://ipsejournal.files.wordpress.com/2015/03/3-ipse-volume-1-no-1-wynne-harlen-p-5-19.pdf

Howes, E. V. (2002). Learning to teach science for all in the elementary grades: What do preservice teachers bring? *Journal of Research in Science Teaching*, *39*(9), 845-869. https://doi.org/10.1002/tea.10049

IBM (2022). KMO and Bartlett's test. Retrieved November 19, 2022, from: https://www.ibm.com/docs/es/spss-statistics/29.0.0?topic=detection-kmo-bartletts-test

Kind, V., & Chan, K. K. (2019). Resolving the amalgam: connecting pedagogical content knowledge, content knowledge and pedagogical knowledge. *International Journal of Science Education*, *41*(7), 964-978. https://doi.org/10.1080/09500693.2019.1584931

Laerd Statistics (2019). Fleiss' kappa using SPSS statistics. Retrieved November 17, 2022, from https://statistics.laerd.com/spss-tutorials/fleiss-kappa-in-spss-statistics.php

Lange, R. T. (2011). Inter-rater reliability. In J. S. Kreutzer, J. DeLuca, & B. Caplan (eds.), *Encyclopedia of Clinical Neuropsychology*. Springer.

Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., & Tomás-Marco, I. (2014). Exploratory item factor analysis: a practical guide revised and updated. *Annals of Psychology*, *30*(3), 1151-1169. http://dx.doi.org/10.6018/analesps.30.3.199361

Mayring, P. (2000). Qualitative content analysis. *Forum: Qualitative Social Research, 1*(2), 1-10. https://doi.org/10.17169/fqs-1.2.1089

Menon, D., & Sadler, T. D. (2016). Preservice elementary teachers' science self-efficacy beliefs and science content knowledge. *Journal of Science Teacher Education*, *27*(6), 649-673. https://doi.org/10.1007/s10972-016-9479-y

Montero-Pau, J., Tierno, S. P., & Tuzón, P. (2017). Are pre-service Primary School teachers prepared to teach science by inquiry? In *3rd International Conference on Higher Education* (pp. 1294-1301). Valencia (Spain). http://dx.doi.org/10.4995/HEAd17.2017.5586

Mullis, I. V. S., & Martin, M. O. (Eds.). (2017). *TIMSS 2019 Assessment Frameworks*. Boston College, TIMSS & PIRLS International Study Center. http://timssandpirls.bc.edu/timss2019/frameworks/

Muñoz, I. A., Resino, D. A., Ruiz-Lázaro, J., Sánchez-Munilla, M., de Miguel, C. R., Constante-Amores, A., & Navarro-Asencio, E. (2022). Perfil de acceso a la universidad de los maestros en España [University access profile of teachers in Spain]. *Educación XX1, 25*(2), 39-63. https://doi.org/10.5944/educxx1.31924

National Research Council [NRC]. (1996). *National science education standards*. National Academies Press.

National Research Council [NRC]. (2012). *A framework for K-12 science education*. National Academies Press.

Newman, W. J., Abell, S. K., Hubbard, P. D., McDonald, J., Otaala, J., & Martini, M. (2004). Dilemmas of teaching inquiry in elementary science methods. *Journal of Science Teacher Education, 15*(4), 257-279. https://doi.org/10.1023/B:JSTE.0000048330.07586.d6

Oh, P. S. (2010). How can teachers help students formulate scientific hypotheses? Some strategies found in abductive inquiry activities of earth science. *International Journal of Science Education, 32*(4), 541-560. https://doi.org/10.1080/09500690903104457

Organisation for Economic Co-operation and Development [OECD]. (1999). *Measuring students' knowledge and skills: a new framework for assessment*. OECD.

Organisation for Economic Co-operation and Development [OECD]. (2019a). *TALIS 2018 Results (Volume I): Teachers and school leaders as lifelong learners*. OECD Publishing.

Organization for Economic Co-Operation and Development [OECD]. (2019b). *PISA 2018 assessment and analytical framework*. OECD Publishing.

Oviedo, H. C., & Campo-Arias, A. (2005). Aproximación al uso del coeficiente alfa de Cronbach [Approximation to the use of Cronbach's alpha coefficient]. *Revista Colombiana de Psiquiatría*, *34*(4), 572-580. https://www.redalyc.org/articulo.oa?id=80634409

Padilla, M. J. (1990). The science process skills. Research matters—To the science teacher, No. 9004. Reston, VA: National Association for Research in Science Teaching (NARST). https://narst.org/research-matters/science-process-skills

Pallant, J. (2011). *SPSS survival manual: A step by step guide to data analysis using SPSS* (4th ed.). Open University Press.

Pellegrino, J. W., DiBello, L. V., & Goldman, S. R. (2016). A framework for conceptualizing and evaluating the validity of instructionally relevant assessments. *Educational Psychologist*, *51*(1), 59-81. https://doi.org/10.1080/00461520.2016.1145550

Rintayati, P., Syawaludin, A., & Kamsiyati, S. (2020). Enhancing elementary teachers' science process skills through participatory scientific approach training. *Journal of Physics: Conference Series,*1511. https://doi.org/0.1088/1742-6596/1511/1/012109

Royal Decree 157/2022, of 1 March, which establishes the organisation and minimum teaching of primary education. *Spanish Official State Gazette, 52*, 2 March 2022, pages 24386 to 24504. https://www.boe.es/eli/es/rd/2022/03/01/157/con

Royal Decree 95/2022, of 1 February, which establishes the organisation and minimum teaching of childhood education. *Spanish Official State Gazette, 28*, 2 February 2022, pages 14561 a 14595. https://www.boe.es/eli/es/rd/2022/02/01/95

Shahali, E. H. M., & Halim, L. (2010). Development and validation of a test of integrated science process skills. *Procedia-Social and Behavioral Sciences*, *9*, 142-146. http: https://doi.org/10.1016/j.sbspro.2010.12.127

Shulman, L. (1986) Those who understand: Knowledge growth in teaching. *Educational Researcher 15*(2), 4-14. https://doi.org/10.3102/0013189X015002004

Šmida, D., Čipková, E., & Fuchs, M. (2023). Developing the test of inquiry skills: measuring the level of inquiry skills among pupils in Slovakia. *International Journal of Science Education*, 1-36. https://doi.org/10.1080/09500693.2023.2219811

Verdugo, J. J., Solaz, J. J., & Sanjosé, V. (2016). Pre-service primary school teachers' science content knowledge: An instrument for its assessment. *International Journal of Innovation in Science and Mathematics Education, 24*(2), 37-51. http://hdl.handle.net/10550/55057

Wilke, R. R., & Straits, W. J. (2005). Practical advice for teaching inquiry-based science process skills in the biological sciences. *The American Biology Teacher*, *67*(9),534-540. https://doi.org/10.2307/4451905

Yoon, H. G., Joung, Y. J., & Kim, M. (2012). The challenges of science inquiry teaching for pre-service teachers in elementary classrooms: Difficulties on and under the scene. *Research in Science Education, 42*(3), 589–608. https://doi.org/10.1007/s11165-011-9212-y

Appendix

Basic science process skills diagnostic questionnaire

(Demographic information items omitted)

1. A science teacher wants to propose several activities to her students for them to develop skills of scientific inquiry. To this end, she formulated the following questions. Which do you think would be conducive to carrying out scientific inquiries at school.

a) What does the size of an object's shadow depend on?
b) Why is Earth the third closest planet to the Sun?
c) How does temperature influence a substance's solubility in water?
d) Why is the Moon a satellite of the Earth?
e) What will happen if we add salt to a piece of ice?
f) Why don't plants have legs?
g) Why do elephants have trunks?
h) Which substance is a better conductor of electricity, sea water or sunflower oil?
i) Why did the dinosaurs become extinct?

2. Pupils in a science class were organized into four groups to inquire into the following question: *"How does the change in mass of a certain material affect its buoyancy in water?"* The hypotheses the groups formulated are indicated below. Indicate which of them you think constitute scientific hypotheses, regardless of whether or not they are valid.

a) What you have to do is change the size of the material, for example, using different pieces of the same metal object, the same wood, or the same liquid (e.g., oil) to see in each case which one floats or sinks more.
b) If the mass of the material increases, its weight will increase. Therefore, its buoyancy will decrease.
c) Take a piece of wood of 100 g and another of 250 g, put them in water, and observe which of them has the greatest sunken portion.
d) If the mass of the material changes, its volume will too in the same proportion, so that the density will stay the same, and consequently neither will its buoyancy change.

3. In the following, various questions are presented with which to initiate school scientific inquiries. Show in each case which is the independent variable (IV), i.e., one that can be manipulated by the researcher, and which the dependent variable (DV), i.e., the one observed and measured as a result of the inquiry.

a) How does the air pressure inside a balloon change with temperature?
IV: ______________________________________
DV: _____________________________________

b) Where will a chickpea seed grow faster, in cotton-wool or in soil?
IV: ______________________________________
DV: _____________________________________

c) Which will conduct electricity better, a thick copper wire or a thin one?
IV: ______________________________________
DV: _____________________________________

d) When allowing a 2 kg box and a 7 kg box fall down a smooth ramp, which one will reach the bottom first?
IV: ______________________________________
DV: _____________________________________

4. A teacher posed his students the following inquiry problem: *"How does the amount of irrigation water affect the growth of a plant?"* Students were asked to describe their methods or strategies for doing the inquiry. These were some of the responses:

Pepe: I will find out the amount of water needed to make the plant grow faster by careful observation.

Laura: I will check if irrigating a plant with water affects its growth during a certain period of time.

Dani: I'll take several plants of the same type and size, add different amounts of water to them over a period of time, and see what happens.

Raquel: I will check how different amounts of water affect the growth of an irrigated plant.

Indicate which of the following explanations you consider best analyses the proposals of the students to address the inquiry problem formulated:

a) The best approach is the one proposed by Dani because it shows that he is clear about the objective and procedure of the investigation. Instead, Pepe's proposal suggests that he has misunderstood what he intends to investigate.
b) The best approach is the one proposed by Pepe because he defines both the objective of the investigation very well and a method to approach it. And the worst, Dani's because he only describes a procedure to carry out the investigation.
c) The best approach is Laura's because it specifically indicates what should be investigated. And the worst, Pepe's because he proposes to find out something that does not respond to the inquiry problem.
d) The best approach is Dani's because he shows that he has understood the inquiry problem and proposes a specific and reliable procedure to address it. On the other hand, the worst approach is Pepe's or Laura's because they do not try to answer the research problem.
e) The best approach is Dani's because she gives more details than her colleagues to carry out an investigation according to the problem formulated.

5. Patricia wants to measure the length of the diagonal of her teacher's rectangular table (the length is 1.7 m). To do this, she has different types of rulers:

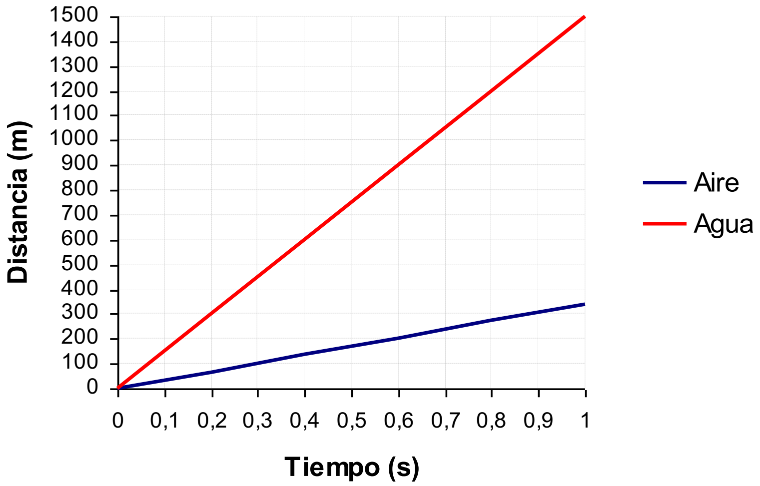
- Ruler of 120 cm, whose minimum unit is 1 mm.
- Ruler of 16 dm, whose minimum unit is 0.1 dm.

- Ruler of 1.5 m, whose minimum unit is 0.5 cm.

If you were her, which of the following rulers available would you use? (Choose one option)

a) The 120 cm ruler because, since none allows you to measure just once, it is the one that will allow greatest precision, and therefore make the smallest error.
b) The 120 cm ruler because with the others you would have less precision, since with none you can make the measurement in one go.
c) The 120 cm ruler or the 16 dm ruler because, in reality, both have a similar precision with which to make the requested measurement.
d) The 16 dm ruler because, being the longest, it is the one that allows you to measure a larger part of the table with a single measurement and has a precision (1 cm) that could be considered acceptable.
e) The 1.5 m ruler because it uses a scale in 'cm', which is the one with which pupils are usually more familiar and that will help them measure better.

6. Sound only propagates through material media, and its speed varies according to the characteristics of the medium. Scientists have measured the distance sound travels as a function of time in air and in seawater. The results are shown in the following graph:



(Distancia: Distance; Tiempo: Time; Aire: Air; Agua: Water)

We know that whales communicate by making very special sounds. A whale is located by some marine biologists 300 m from the shore of a beach just when it has come to the surface. At that moment, it emits one of its characteristic sounds.

According to the graph above, who would receive that sound first, the biologists located on the shore (they have special equipment for capturing sounds on the surface) or another whale that is under the sea 800 m from the first? (Choose an option):

a) The biologists will receive it earlier because they have a more sophisticated sound detector than whales.
b) The whale under the sea will receive it sooner, in approximately half the time it will take to reach the biologists.
c) The biologists will receive it sooner because sound propagates more slowly in the water.
d) The whale will receive it earlier under the sea because the speed of sound in this medium is greater than in air.
e) The whale will receive it earlier under the sea, in somewhat more than 0.5 s.

7. In a school laboratory, Laura and Alfonso have measured the solubility of potassium iodide in water at different temperatures. The data is given in the following table:

Temperature (°C)	Solubility of potassium iodide (grams in 100 g of water)
20	145
40	160
60	175
80	190
100	210

Imagine that 300 g of potassium iodide are added to 200 g of water at 50°C. According to the table above, what will happen? (Choose the most appropriate option):

a) It will dissolve completely because, at 60°C, it is possible to dissolve 175 g of potassium iodide for every 100 g of water.
b) Not all of it will dissolve because, even at double the temperature, only 210 g of potassium iodide can be dissolved.
c) It will dissolve completely because it is equivalent to dissolving 150 g of potassium iodide in 100 g of water, and this is achieved at around 40°C.
d) It will not dissolve completely because the amount of potassium iodide (solute) is greater than that of water (solvent).
e) It will dissolve completely because, doubling the amount of water favours the solubility of potassium iodide.
f) Not all of it will dissolve because, at 50°C, only the amount of potassium iodide can be dissolved that a temperature of between 40° and 60°C allows.