Psychol Stud (January–March 2025) 70(1):148–158
https://doi.org/10.1007/s12646-025-00819-4

ASSESSMENT

Spanish-Speaking Children's Attitudes Toward School Science: Instrument Development and Psychometric Analysis

César Quílez-Cervero[1] · Radu Bogdan Toma[1] · Miguel Ángel Queiruga-Dios[1]

Received: 28 February 2023 / Accepted: 9 January 2025 / Published online: 31 January 2025

Abstract The current research evaluated the psychometric properties of a questionnaire measuring 1st and 2nd graders' attitudes toward school science (CASS). In Study 1, an exploratory factor analysis revealed a two-factor structure, consisting of the 'enjoyableness' and 'self-efficacy' dimensions of school science. This analysis also provided initial evidence of satisfactory internal consistency reliability. In Study 2, confirmatory factor analysis showed that the two-factor model provided a better fit than a unidimensional structure and further supported the internal consistency reliability of the instrument. Study 3 provided strong evidence for test–retest reliability. Study 4 offered promising evidence of construct validity by identifying a positive and significant relationship between the 'enjoyableness' and 'perceived self-efficacy' in school science and leisure interest in science-related activities. Finally, Study 5 assessed the instrument's responsiveness to an inquiry-based intervention, showing that it can effectively detect changes over time. Taken together, these findings suggest that the proposed instrument, CASS, yields scores with adequate validity and reliability, making it a reliable tool for assessing young students' attitudes toward school science and evaluating the effectiveness of educational interventions in the early elementary school years.

Keywords Attitudes toward science · Elementary grades · Children · Self-efficacy · Enjoyableness

Introduction

The decline in positive attitudes toward science has prompted the development of major educational efforts to repair the leakages in the science pipeline (Jiang & McComas, 2015; for a review, see van den Hurk et al., 2019). Despite the efforts, no significant improvements have been made. Students are becoming increasingly disengaged from science, and research suggests that by the end of elementary grades, it may be too late to stimulate their scientific vocations (Maltese & Tai, 2011; Toma, 2024). Worldwide research studies undertaken with students in late elementary and early secondary school demonstrate a consistent trend: Students' attitudes toward science decline as the academic year increases (Ali et al., 2013; Denessen et al., 2015; DeWitt & Archer, 2015; Said et al., 2016).

This scenario is alarming across the world, especially when considering the importance of scientific literacy for making informed decisions, regardless of whether students choose to pursue a science-related career or not. In other words, positive attitudes toward science are also required for making educated judgments and decisions on contemporaneous socio-scientific issues (Roberts & Bybee, 2014). As a result, there is a call for increased attention to lower elementary grades (Miller, 2021). Such an endeavor, however, has been hampered by the lack of appropriate and easy-to-use measurement instruments for the assessment of attitudes toward science in early grades (Blalock et al., 2008; Toma, 2021a, 2021b). Consequently, the purpose of this study is to address this gap in the literature by describing the design and psychometric evaluation of short, valid, and reliable instruments that can be used with confidence in the 1st and 2nd grades of elementary education.

✉ Miguel Ángel Queiruga-Dios
maqueiruga@ubu.es

[1] Department of Specific Didactics, Faculty of Education, University of Burgos, C/Villadiego, S/N, 09001 Burgos, Spain

Theoretical Underpinnings

"Attitudes toward science" is a polysemous construct, often conflated with interest or motivation. In the literature, there is a general agreement in considering attitudes as "a psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor" (Eagly & Mladinic, 1993, p. 1). Hence, attitudes are a multidimensional construct composed of cognitive (thoughts and beliefs), affective (feelings and emotions), and behavioral (past and future activities and experiences) components (Breckler, 1984; Potvin & Hasni, 2014; Toma & Lederman, 2022; Tytler, 2014).

In science education research, a distinction is needed between "scientific attitudes" and "attitudes toward science." The former refers to those characteristics common to scientists and scientific endeavors, such as curiosity, objectivity, or critical-mindedness (Gardner, 1975). The latter is to the evaluative judgments of different scientific-related attitudinal objects, such as science as an enterprise, scientific research, scientists, or school science, among many others (Tytler, 2014; Tytler & Osborne, 2012). In this sense, Koballa and Crawley (1985) defined attitudes toward science as an enduring positive or negative feeling about science nearly four decades ago. However, there is still a lot of ambiguity today (Potvin & Hasni, 2014).

Uncertainty increases when attitudes toward science instruments are reviewed. In this sense, researchers have used many different dimensions thought to be measuring attitudes toward science. Examples include, but are not limited to anxiety, career interest, normality of scientists, control beliefs, or emotional satisfaction (Bauer, 2008; Navarro et al., 2016; Summers & Abd-El-Khalick, 2018). In addition, there are also concerns about the overall quality of existing measurement instruments. Blalock et al. (2008) warned that many attitudinal instruments lacked the most basic evidence of validity and reliability to be able to recommend their use. Years later, Toma and Lederman (2022) reviewed attitudinal instruments from 2008 to the present day, reaching very similar conclusions. Thus, most of the instruments had undergone little psychometric analysis and, some of them widely used in several studies (e.g., ROSE), obtained a contradictory latent dimensionality (Toma, 2021a, 2021b).

These same literature review studies have identified a gap in instruments for the initial years of elementary school. Except for the Leisure Time in Science (LeTiS) questionnaire (Paños & Ruiz-Gallardo, 2021), which was developed for the beginning years of elementary school, the majority of the instruments for this educational stage are focused on 4 to 6th graders (for reviews, see Blalock et al., 2008; Toma & Lederman, 2022). As mentioned, the exception to this pattern is the recently published LeTiS instruments. Although it is a conceptually sound measure producing scores with substantial evidence of validity and reliability, its application is limited to science during leisure time, hence not being useful to assess the effectiveness of educational intervention performed during school science classes. In this sense, many investigations reveal that while students appreciate science as a discipline, they reject school science (for reviews, see Osborne et al., 2003; Tytler, 2014). In other words, there is a need for an instrument focused on measuring the attitude toward school science learning, which is the attitudinal object of concern. This measure would make it possible to evaluate the effectiveness of pedagogical changes in school science and their possible impact on the improvement of lower-elementary graders' attitudes.

The Present Research

The current paper describes the psychometric evaluation of a brief questionnaire that assesses the attitudes of 1st and 2nd-grade students toward school science, which was called CASS (children's attitude toward school science). This is an instrumental research project that includes studies on the psychometric properties of a new measurement instrument (Ato et al., 2013). The significance of this research project consists in filling the gap for valid and reliable attitude measures for elementary school students in lower grades.

To this end, the four-step procedure proposed by Netemeyer et al. (2003) was followed. During the first step (construct definition), two authors of this study, with extensive experience in attitudinal and psychometric research, reviewed the established literature on attitudes and defined the construct under study. Two overarching constructs were chosen (i.e., enjoyableness and self-efficacy) since these are some of the attitudinal dimensions most investigated (Kennedy et al., 2016; Toma & Lederman, 2022).

In the second step (generating measurement items), the authors developed an initial pool of 20 items which were examined for clarity by a panel of 27 elementary school teachers. Suggestions included adapting the language to enhance comprehension and dropping long and difficult items, such as "School science lessons are easier for me than for my peers" or "My school science teacher considers that I do well in school science." This procedure led to the final pool of 13 items that were administered to a large-scale sample for psychometric analysis. The panel also recommended using a written response format (i.e., yes, so-so; no) rather than an emoji response scale in gray; an emoji response scale with red, yellow, and green colors; and a thumbs up–thumbs down response format.

Following the reputable Standards for educational and psychological testing (AERA et al., 2014), and the international COSMIN taxonomy of measurement properties (Mokkink et al., 2010), the third (designing and conducting studies to refine the scale) and fourth step (finalizing the scale)

consisted of assessing the instrument against structural, and hypothesis-testing validity; internal consistency and temporal stability (test–retest); and responsiveness.

Study 1

The purpose of Study 1 was to examine the latent structure of the proposed instrument using robust exploratory factor analysis, which provides evidence of structural validity, i.e., the degree to which the scores are an adequate reflection of the dimensionality of the attitude construct (Mokkink et al., 2010, p. 743). Likewise, evidence for internal consistency reliability was gathered, i.e., the degree of the interrelatedness among the items of a construct (Mokkink et al., 2010, p. 743).

Participants

According to sample size recommendations for exploratory factor analysis, an instrument with at least 6 variables per factor and wide commonalities (0.20 to 0.80), a sample size of 200 participants would result in robust solutions (Gaskin & Happell, 2014). Hence, a total of 300 students from state-funded (63.3%) and semi-private (36.7%) schools were surveyed. The sample was drawn using convenience sampling techniques from the urban area of Burgos, a city located in north-central Spain (Cohen et al., 2018), yet an effort was made to represent the educational milieu of such a context. Almost half of the sample were girls (46%), and students were aged 6 to 8 years old ($M = 6.94$, SD = 0.68).

Measures

The proposed questionnaire consisted of 13 items (Table 1). Seven items were developed to measure the enjoyableness of school science, and the remaining six items were intended to measure self-efficacy.

Data Analysis

Responses were subjected to robust exploratory factor analysis (EFA) using the maximum likelihood extraction procedure with oblique Promax rotation, as this procedure provides more accurate results when constructs are conceptually related (Gaskin & Happell, 2014; Roberson et al., 2014). Up to three different factor retention criteria were used: Cattell's (1966) scree test, parallel analysis, and minimum average partial (MAP) test (O'Connor, 2000). Items with loadings below 0.32 or cross-loadings between factors were a candidate for removal (Tabachnick & Fidell, 2007).

Regarding evidence of reliability, the Cronbach alpha (α) and McDonald's omega (ω) were used as means of internal consistency reliability, the latter being more suitable for Likert-type items (Hayes & Coutts, 2020; Peters, 2014).

All analyses were performed using SPSS v.25 (IBM, 2017), with the OMEGA macro for calculating McDonald's omega (Hayes & Coutts, 2020).

Table 1 The initial pool of items

Construct	Items
Enjoyableness	School science lessons are…
	i. Fun
	ii. Of my interest
	iii. My favorites
	iv. I like them
	In school science lessons…
	v. I am very attentive
	vi. I learn interesting things
	vii. I have a good time
Self-efficacy	School science lessons are…
	i. Difficult
	ii. Complicated
	In school science lessons…
	iii. I make a lot of mistakes
	iv. I learn things quickly
	v. I know how to do it well
	vi. I need help

It should be noted that the items were written in Spanish; therefore, their translation into English is for reporting purposes and does not have the same grammatical consistency

Results

The Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy was 78 and passed Bartlett's test of sphericity ($\chi^2(78) = 1152.80$; $p < 0.01$), which confirms the suitability of the dataset for factor analysis (Tabachnick & Fidell, 2007). Cattell's scree test indicated that a two-factor structure suits best the data (Fig. 1). The decision to extract two factors was further supported by the results of parallel analysis and MAP (O'Connor, 2000), suggesting a two-factor latent structure.

The factor analysis with the maximum likelihood extraction method and oblique Promax rotation revealed that two enjoyableness items (I am very attentive; I learn interesting things) had low factor loadings (< 0.32). After the removal of these items, a parsimonious structure was achieved, consisting of two constructs explaining a total of 50.80% of the item variance (Table 2). The first factor has an eigenvalue of 3.63 and is composed of 5 items measuring the "enjoyableness of school science." The second factor has an eigenvalue of 1.96 and consists of 6 items measuring "self-efficacy in school science." These findings provide initial evidence for structural validity.

Next, for the sake of parsimony and simplicity, the length of the scales was reduced based on the following criteria: (i) high factor loadings; (ii) different wording, thus representing

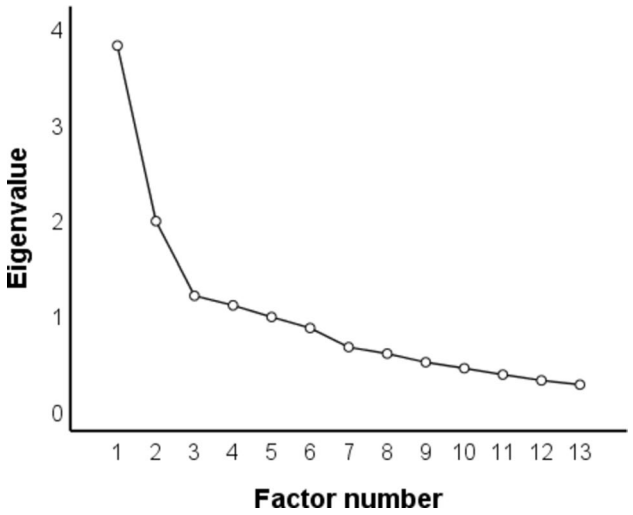

Fig. 1 Cattell's scree test

Table 2 Results of exploratory factor analysis

	Factors	
	1	2
Factor 1: Enjoyableness of school science		
Fun	.84	
I like them	.80	
I have a good time	.76	
Of my interest	.49	
My favorites	.48	
Factor 2: Self-efficacy in school science		
Difficult		.80
Complicated		.72
I need help		.65
I make a lot of mistakes		.49
I know how to do it well		.36
I learn things quickly		.32

construct breadth; (iii) maintaining adequate reliabilities indices (> 0.60); and (iv) minimum of 3 items per construct (DeVellis, 2017). Exploratory factor analysis and reliability results of the reduced version are depicted in Table 3. The reduced version explained a total of 68.52% of the variance and both factors had high eigenvalues (2.57 and 1.54, respectively) and great factor loadings. Likewise, internal consistency reliability was adequate.

Summary

Study 1 examined the dimensionality of the proposed instrument. Overall, findings suggest that enjoyableness and self-efficacy in school science are two distinct constructs that can be measured with only three items while maintaining adequate reliability, particularly for scales with few items and aimed at young respondents (Newman & McNeil, 1998; Nunnally & Bernstein, 1994).

Table 3 The reduced version of the scale

	Factors	
	1	2
Factor 1: Enjoyableness of school science		
I like them	.84	
Fun	.80	
I have a good time	.76	
Factor 2: Self-efficacy in school science		
I need help		.80
I make a lot of mistakes		.72
Difficult		.65
Cronbach α	.83	.67
McDonald ω	.84	.67

Study 2

The purpose of Study 2 was to assess the model fit of the two-factor structure resulting from Study 1, thus providing further evidence for structural validity in a new sample of students. Likewise, internal consistency reliability was calculated.

Participants

Sample size recommendations for confirmatory factor analysis include at least 200 responses and a 7-to-1 participant-to-item ratio (Kline, 2005; Mundform et al., 2005). Therefore, the sample for this study comprised 282 students from the urban area of Burgos. Of them, 163 and 119 were drawn from state-funded and semi-private schools, respectively, using convenience-sampling techniques. Students (47.2% of girls) were aged 6–9, with a mean age of 6.88 (SD = 0.70).

Measures

The six-item attitudinal instrument from Study 1 was used. The items of each factor were presented in an alternating manner: items 1, 3, and 5 (I like them; Fun; I have a good time) represented enjoyableness. Items 2, 4, and 6 (I need help; I make a lot of mistakes; Difficult) represented self-efficacy.

Data Analysis

Responses were subjected to confirmatory factor analysis (CFA) with the maximum likelihood estimation method to examine the model fit of two theoretical models. The first model was composed of a unidimensional factor measuring overall attitude toward school science. The second model was based on the EFA results from Study 1 and consisted of two, first-order factors measuring enjoyableness, and self-efficacy in school. Each model was tested against three goodness-of-fit indices (Kline, 2005): (i) comparative fit index (CFI) ≥ 0.90; (ii) Tucker–Lewis index (TLI) ≥ 0.90; and (iii) root mean square error of approximation (RMSEA) ≤ 0.08. All analyses were performed using SPSS v.25 and AMOS v.23 software (Arbuckle, 2014; IBM, 2017), with the OMEGA macro for calculating McDonald's omega (Hayes & Coutts, 2020).

Results

The unidimensional latent structure had a poor model fit (CFI = 0.76; TLI = 0.59; RMSEA = 0.20) and loadings below the minimum 0.32 cutoff (Fig. 2). Yet, the two-factor model had adequate model fit (CFI = 0.98; TLI = 0.96; RMSEA = 0.06) and high factor loadings (Fig. 3). Hence, the unidimensional structure was dropped from further analyses.

The reliability of the two-factor structure was satisfactory. Both Cronbach α and McDonald ω were 0.79 and 0.66 for the enjoyableness and self-efficacy factors, respectively.

Summary

Study 2 confirmed the latent structure resulting from EFA analyses in Study 1. Hence, the proposed instrument is consistent with theoretical expectations that attitudes toward school science are a multidimensional construct composed of affective (enjoyableness) and cognitive (self-efficacy) dimensions. Taken together, Study 2 secured further evidence for the structural validity and internal consistency reliability of both scales in a new sample of students.

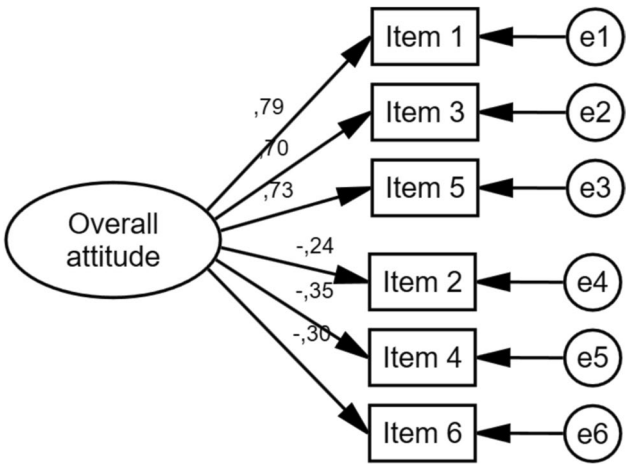

Fig. 2 Standardized estimates for the unidimensional model

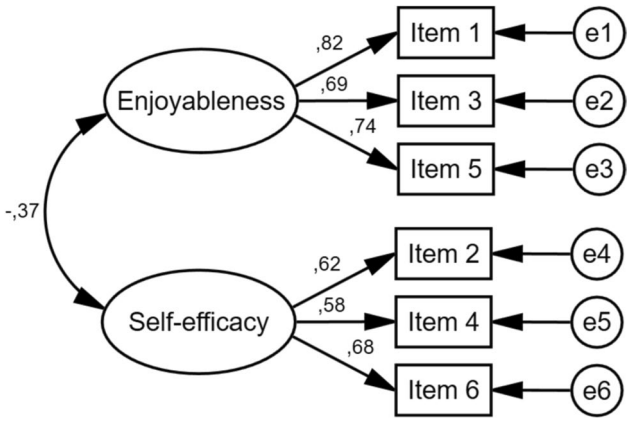

Fig. 3 Standardized estimates for the two-factor model

Study 3

Study 3 aimed at gathering evidence for the temporal stability (test–retest) reliability of the proposed instrument, i.e., the extent to which scores for students who have not changed are the same for repeated measurement over time (Mokkink et al., 2010, p. 743). Such psychometric evidence would be useful when using the instrument in studies with time-series designs and several data collection phases (Toma & Lederman, 2022).

Participants

A convenience sample was drawn from state-funded schools. A total of 107 students (45.8% girls), aged 6–8 ($M = 6.78$; SD = 0.69), were surveyed on two occasions.

Measures and Procedure

The proposed six-item scale was used as described in Study 2. A ten-day span was used between the first and second administration of the scale, as it is considered adequate for computing test–retest reliability (Keszei et al., 2010). Attitudes toward school science can change significantly over time due to various factors. A shorter interval between tests, sufficient to minimize recall effects, mitigates the influence of these external factors. This approach leads to more stable assessments. The close timeframe allows researchers to

obtain consistent measurements that reflect the short-term stability of attitudes.

Data Analysis

ICC estimates and their 95% confidence intervals were calculated based on a mean-rating, absolute-agreement, 2-way mixed-effects model (Koo & Li, 2016). All analyses were performed using SPSS v.25 (IBM, 2017),

Results

ICC estimates and their 95% confidence intervals were 0.83 (0.76–0.89) for the "enjoyableness" scale and 0.76 (0.65–0.84) for the "self-efficacy" scale.

Summary

Study 3 used a convenience sample of students who responded to the questionnaire a second time ten days later and gathered evidence for test–retest reliability. Specifically, ICC estimates and their 95% confidence intervals were indicative of moderate to good test–retest reliability of the proposed instrument (Koo & Li, 2016).

Taken together, the instrument produces reliable scores when used in studies with time-series designs, allowing longitudinal studies to be conducted to unravel when leakage in the scientific pipeline begins.

Study 4

The aim of Study 4 was to assess the construct validity of the proposed instrument, i.e., the degree to which its scores are related to scores of other instruments of conceptual convergence (Mokkink et al., 2010). Thus, the relationship between the enjoyableness and self-efficacy scales and two established questionnaires measuring leisure interest and attitude toward informal science activities was examined. The construct "leisure interest in science" was used as a benchmark to assess the concurrent validity of the CASS instrument for two main reasons. First, students who enjoy science and find it less difficult are more likely to engage with it voluntarily during their free time, indicating that positive attitudes toward school science might extend beyond the classroom. Second, valid and reliable attitudinal instruments for young Spanish-speaking students are limited. The LeTiS Instrument (Paños & Ruiz-Gallardo, 2021), with strong evidence of validity and reliability, provides an ideal standard for comparison.

Participants

The required sample size was determined for a two-tailed Pearson correlation coefficient using G*Power 3 (Faul et al., 2007). For alpha = 0.05 and power = 80%, a large ($r = 0.50$) and small ($r < 0.30$) would require between 29 and 91 participants. Hence, a total of 93 students (46.2% girls) drawn using convenience sampling from state-funded schools comprised the sample. Participants' ages ranged from 6 to 8 years old ($M = 6.81$, SD = 0.71).

Measures

Attitudes Toward School Science

The proposed six-item scale, measuring enjoyableness and self-efficacy, was used as described in Studies 2 and 3.

Leisure Time in Science (LeTiS)

The LeTiS pictographic scale was used to measure students' attitudes toward science learning activities in informal environments (Paños & Ruiz-Gallardo, 2021). This questionnaire consists of eight pictures representing science-related activities (e.g., looking at insects; reading books about animals and plants; and a trip to a zoo and science museum). The scale was designed for Spanish students aged 5–8 and was found to produce results with robust validity and reliability. It was administered using a three-point response option (thumb up = 3 points; thumb sideways = 2 points; thumb down = 1).

Data analysis

Items of the enjoyableness and self-efficacy measures were summed and divided by the total number of items in each scale, thus obtaining values from 1 (minimum) to 5 (maximum). The same procedure was followed for the leisure time in the science questionnaire. Next, the relationship between all three scores was assessed using the Pearson correlation coefficient, since the data passed assumption criteria (Knapp, 2018).

Results

Both the enjoyableness of school science ($r = 0.21$, $p < 0.05$) and self-efficacy ($r = 0.24$, $p < 0.05$) were significantly and positively correlated with attitudes toward science learning activities in informal environments.

Summary

Study 4 gathered construct evidence for the proposed attitude toward school science instruments. Consistent with the hypothesis, scores of the enjoyableness and self-efficacy scales were positively related to scores of a valid and reliable instrument measuring leisure interest in science-related activities. While these results are in line with theoretical expectations, it should be noted that both correlation coefficients were small. Hence, while promising, the results of concurrent validity are not conclusive and warrant further studies.

Study 5

Study 5 aimed at assessing the responsiveness of the enjoyableness and self-efficacy scales, which is defined as the ability of the instrument to detect change over time in the construct to be measured (Mokkink et al., 2010, p. 743). Specifically, participants from Study 4 participated in an inquiry-based educational intervention. Hence, this study used a quasi-experimental, one-group, pretest–posttest design (Shadish et al., 2002).

Inquiry teaching methodologies are well known for improving students' attitudes toward science (Aguilera & Perales-Palacios, 2020; Demirel & Dağyar, 2016). Indeed, a recent meta-analysis found a medium effect size for inquiry-based teaching strategies (Aguilera & Perales-Palacios, 2020). Therefore, it was hypothesized that the intervention would improve students' enjoyment of school science and their self-efficacy beliefs; if the proposed measurement instrument can identify such a change in score, it is assumed to be responsive to interventions.

Participants

The sample consisted of the same participants from Study 5. Yet, only 91 completed questionnaires were gathered (2 students failed to provide post-test data). The required sample size was determined for the related sample t test using G*Power 3 (Faul et al., 2007). For alpha = 0.05 and power = 80%, a medium ($d = 0.5$) effect size, as reported in previous meta-analyses of the literature (Aguilera & Perales-Palacios, 2020), would require between 35 students. Hence, Study 5 has adequate power.

Intervention

A short-term intervention was designed in line with the Spanish curricula and the educational milieu. An 8-h unit addressed the curricular content of minerals and rocks by enacting a structured inquiry strategy wherein students investigated teacher-supplied research questions following a prescribed procedure but not knowing the results in advance (Banchi & Bell, 2008; Yánez-Pérez et al., 2024). Hence, students engaged in different activities such as observation, identification, description, and classification of materials based on their elementary properties. In doing so, the unit tackled the standard from the Spanish curricula named "initiation to scientific activity," which demands the use of inquiry in all the courses of the elementary stage (LOMCE, 2013). Minerals and rocks were also highlighted for their application and utility in everyday life. The didactic unit concluded with an application activity that included a field trip to identify minerals and rocks concerning the town of Burgos' city cultural heritage (e.g., churches, cathedrals, historic buildings).

Measures

The proposed six-item scale, measuring enjoyableness and self-efficacy, was used as described in Studies 2–4.

Data Analysis

Students' responses were analyzed using a paired sample t test after confirming that the data met the assumption criteria (Knapp, 2018). This test focuses on the differences between paired observations, such as posttest scores minus pretest scores. A critical assumption of the t test is that these differences are normally distributed. In this study, this assumption was satisfied, as the skewness and kurtosis values fell within the ± 3 cutoff range. A homogeneity of variance test was unnecessary because the analysis involved the same group.

Results

Figure 4 reports the mean value for pretest and posttest scores. Before the intervention, enjoyableness ($M = 2.67$; SD = 0.58) and self-efficacy ($M = 2.35$; SD = 0.60) scores were high. Following the intervention, both enjoyableness ($M = 2.72$; SD = 0.52) and self-efficacy ($M = 2.48$; SD = 0.51) scores improved. A paired sample t test revealed that the improvement for the enjoyableness scale was negligible, $t(90) = -1.14$, $p = 0.27$. Yet, changes in self-efficacy score were statistically significant, $t(90) = -2.56$, $p = 0.01$.

Summary

The purpose of Study 5 was to examine if the proposed instrument could detect change over time. For this purpose, students were subjected to an inquiry-based unit to improve their attitudes toward school science (Aguilera & Perales-Palacios, 2020). Overall, the proposed instrument revealed

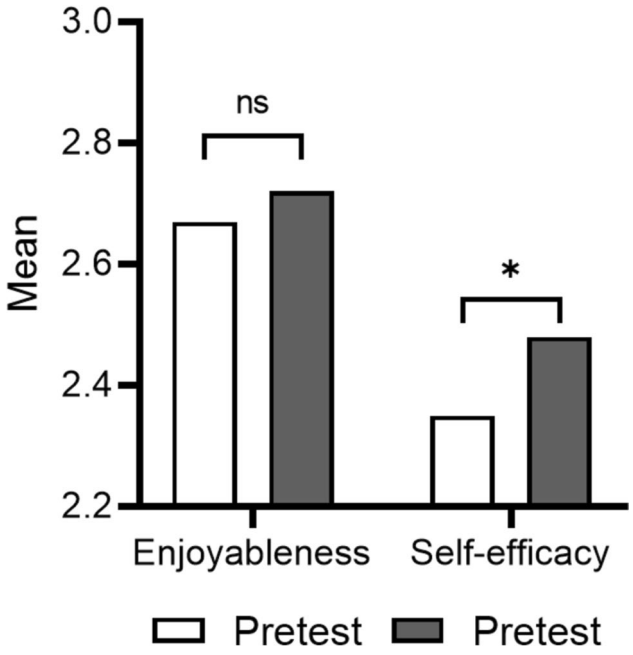

Fig. 4 Pretest and posttest means. Note: ns = non-significant ($p > .05$); * = significant ($p < .05$)

that students' enjoyment and self-efficacy beliefs improved. However, only the self-efficacy scores showed a statistically significant improvement. The lack of statistically significant changes in enjoyment could be explained by the ceiling effect. Before the intervention, the students in this sample had a high overall enjoyment score.

General Discussion and Conclusions

Many studies in recent years have found a decrease in students' attitudes toward science as school grades increase (Aschbacher et al., 2014; DeWitt & Archer, 2015; Kennedy et al., 2014). The "leakages in the science pipeline" metaphor is frequently used to explain this phenomenon. Despite the availability of numerous attitudinal measurement instruments, there was a scarcity of psychometrically sound measures for the lower grades of elementary education (Blalock et al., 2008; Toma & Lederman, 2022). Improving young students' attitudes is critical for the development of successful educational measures aimed at repairing pipeline leaks (Miller, 2021).

Drawing on the tri-partite model of attitudes (Breckler, 1984) and building upon existing measures focusing on the affective and cognitive components of attitudes (e.g., Xu & Lewis, 2011; Zhang & Campbell, 2011), this research project designed a questionnaire that produces valid and reliable scores for 1st and 2nd elementary graders enjoyableness of school science and self-efficacy beliefs in school science. In line with the Standards for educational and psychological testing (AERA et al., 2014) and the measurement properties taxonomy of the COSMIN initiative (Mokkink et al., 2010), the psychometric investigations conducted in this research project indicated a two-factor latent structure with a satisfactory model fit, indicating adequate structural validity. Similarly, scores were associated with an existing measure of conceptual convergence, indicating that the construct validity of the test was promising. Internal consistency coefficients for short scales directed at young respondents were within acceptable limits in the reliability domain. Moreover, the instruments exhibited good temporal stability (test–retest), making it a good choice for research involving several data collection phases, i.e., longitudinal designs or intervention studies with interrupted time-series designs (Shadish et al., 2002). Finally, the proposed scales appear to be responsive and capable of detecting changes in students' scores following an educational intervention, particularly in the domain of self-efficacy.

Implications

This work, taken together, results in a short questionnaire with robust psychometric properties for measuring 1st and 2nd elementary graders' attitudes toward school science. Each dimension is measured with only three items and three response options, making it appropriate for pupils aged 6 to 9. Because of its simplicity, the instrument can be utilized without causing undue stress to teachers, increasing their willingness to engage in research projects. It can also be used in conjunction with other metrics to determine what factors influence young children's favorable attitudes toward science in school.

Limitations and Avenues for Future Research

Several study limitations must be acknowledged. First, the instrument is based on the tripartite model of attitudes, which considers affective, cognitive, and behavioral elements in the formation of attitudes. However, only two scales that measure affective and cognitive domains have been proposed and tested against current validity and reliability findings. This was done to align with other attitudinal instruments that concentrate on such dimensions. An instrument that measures behavioral intention to enroll/study school science, on the other hand, would enable predictive and correlational studies to disentangle the relationship between the three components of attitudes.

Second, while structural validity and internal consistency reliability (Studies 1 and 2) were collected for students from both state-funded and semi-private schools due to convenience sampling, the test–retest reliability, the construct validity, and the responsiveness of the

questionnaire were only examined for students from state-funded schools. While semi-private schools in Spain are likewise state-funded (but privately owned), and student demographics are similar in both state-funded and semi-private schools, further research on semi-private children is needed to generalize the findings.

Third, the instrument's construct validity is promising, but we expected stronger links between enjoyment, self-efficacy, and leisure interest in science-related activities. There is a need for more research on the relationship between the designed questionnaire and other existing measures of conceptual convergence. However, the dearth of reliable measurement instruments for young children prevented such an approach. Indeed, a series of literature reviews of attitudinal instruments found that, overall, existing instruments had not been validated against the essential validity and reliability properties to be used with confidence (Blalock et al., 2008; Toma & Lederman, 2022).

Finally, the construct validity of the instrument is promising, but we expected higher correlations between enjoyableness, self-efficacy, and leisure interest in science-related activities. Additional studies further addressing the relationship between the proposed questionnaire and other existing measures of conceptual convergence are needed. Yet, such an approach was hampered by the lack of existing robust measurement instruments for young students. Indeed, a series of literature reviews on measurement properties of attitudinal instruments concluded that, overall, existing instruments were not assessed against the necessary validity and reliability properties to use them with confidence (Blalock et al., 2008; Toma & Lederman, 2022). This is an important caveat in Spanish-related literature. Toma (2021a, 2021b) conducted a review of the most widely used attitude instrument in Spain, determining that the evidence of validity and reliability found in the literature was insufficient and well below current requirements.

While the CASS instrument offers valuable insights into students' affective and cognitive attitudes toward science, addressing these limitations through future research and instrument refinement would enhance its predictive power, generalizability, and overall validity. Despite these limitations, the proposed instrument is a promising tool for filling the gap in valid and reliable attitudes toward science instruments in early primary school.

Acknowledgements This study was supported by the Spanish Ministry of Economy, Industry, and Competitiveness (MINECO) through the project PID2020-117348RB-I00. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions Conceptualization, supervision, validation and writing —review and editing were done by RBT and MAQ; data curation and formal analysis were done by RBT; funding acquisition was done by MAQ; investigation, methodology, project administration, resources, and visualization were done by CQC, RBT, and MAQ; and writing—original draft was done by CQC.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This study was supported by the Spanish Ministry of Economy, Industry and Competitiveness (MINECO) through the project PID2020-117348RB-I00. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request. These data are currently being used as part of the first author's Ph.D. dissertation. Once he defends his dissertation, the data will be made publicly available in a repository of the University of Burgos.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Consent to Participate Informed consent was obtained from participants and their parents or legal guardians.

Consent to Publish Informed consent was obtained to publish the data generated in this study at group level and after being anonymized.

Ethical Approval The study was approved by the Vicerectorate of Research and the Ph.D. program in education at the University of Burgos. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

References

AERA, NCME, & APA. (2014). *Standards for educational and psychological testing*. American Psychological Association.

Aguilera, D., & Perales-Palacios, F. J. (2020). What effects do didactic interventions have on students' attitudes towards science? *A Meta-Analysis. Research in Science Education, 50*(2), 573–597. https://doi.org/10.1007/s11165-018-9702-2

Ali, M. M., Yager, R., Hacieminoglu, E., & Caliskan, I. (2013). Changes in student attitudes regarding science when taught by teachers without experiences with a model professional development program. *School Science and Mathematics, 113*(3), 109–119. https://doi.org/10.1111/ssm.12008

Arbuckle, J. L. (2014). *Amos (Version 23.0)*. IBM SPSS.

Aschbacher, P. R., Ing, M., & Tsai, S. M. (2014). Is Science me? Exploring middle school students' STE-M career aspirations. *Journal of Science Education and Technology, 23*(6), 735–743. https://doi.org/10.1007/s10956-014-9504-x

Ato, M., López, J. J., & Benavente, A. (2013). A classification system for research designs in psychology. *Anales De Psicologia, 29*(3), 1038–1059. https://doi.org/10.6018/analesps.29.3.178511

Banchi, H., & Bell, R. (2008). The many levels of inquiry. *Science and Children, 46*(2), 26–29.

Bauer, C. F. (2008). Attitude toward chemistry: A semantic differential instrument for assessing curriculum impacts. *Journal of Chemical Education, 85*(10), 1440–1445. https://doi.org/10.1021/ed085p1440

Blalock, C. L., Lichtenstein, M. J., Owen, S., Pruski, L., Marshall, C., & Toepperwein, M. A. (2008). In pursuit of validity: A comprehensive review of science attitude instruments 1935–2005. *International Journal of Science Education, 30*(7), 961–977. https://doi.org/10.1080/09500690701344578

Breckler, S. J. (1984). Empirical validation of affect, behavior, and cognition as distinct components of attitude. *Journal of Personality and Social Psychology, 47*(6), 1191–1205. https://doi.org/10.1037/0022-3514.47.6.1191

Catell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*, 245–276.

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education (8th edition)*. Routledge.

Demirel, M., & Dağyar, M. (2016). Effects of problem-based learning on attitude: A meta-analysis study Eurasia Journal of Mathematics. *Science and Technology Education, 12*(8), 2115–2137. https://doi.org/10.12973/eurasia.2016.1293a

Denessen, E., Vos, N., Hasselman, F., Louws, M., Denessen, E., Vos, N., Hasselman, F., & Louws, M. (2015). The relationship between primary school teacher and student attitudes towards science and technology. *Education Research International, 2015*, 1–7. https://doi.org/10.1155/2015/534690

DeVellis, R. F. (2017). *Scale development: Theory and applications*. Sage Publications Inc.

DeWitt, J., & Archer, L. (2015). Who aspires to a science career? A comparison of survey responses from primary and secondary school students. *International Journal of Science Education, 37*(13), 2170–2192. https://doi.org/10.1080/09500693.2015.1071899

Eagly, A. H., & Mladinic, A. (1993). Are people prejudiced against women? Some answers from research on attitudes, gender stereotypes, and judgments of competence. *European Review of Social Psychology, 5*(1), 1–35.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146

Gardner, P. L. (1975). Attitudes to science: A review. *Studies in Science Education, 2*, 1–41.

Gaskin, C. J., & Happell, B. (2014). On exploratory factor analysis: A review of recent evidence, an assessment of current practice, and recommendations for future use. *International Journal of Nursing Studies, 51*, 511–521. https://doi.org/10.1016/j.ijnurstu.2013.10.005

Hayes, A. F., & Coutts, J. J. (2020). Use Omega rather than Cronbach's Alpha for estimating reliability. *Communication Methods and Measures, 14*(1), 1–24. https://doi.org/10.1080/19312458.2020.1718629

IBM, C. (2017). *IBM SPSS Statistics for Windows. Version 25.0*. IBM Corp.

Jiang, F., & McComas, W. F. (2015). The effects of inquiry teaching on student science achievement and attitudes: Evidence from propensity score analysis of PISA data. *International Journal of Science Education, 37*(3), 554–576. https://doi.org/10.1080/09500693.2014.1000426

Kennedy, J., Lyons, T., & Quinn, F. (2014). The continuing decline os science and mathematics enrolments in Australian high schools. *Teaching Science, 60*(2), 34–46.

Kennedy, J., Quinn, F., & Taylor, N. (2016). The school science attitude survey: A new instrument for measuring attitudes towards school science. *International Journal of Research & Method in Education, 39*(4), 422–445. https://doi.org/10.1080/1743727X.2016.1160046

Keszei, A. P., Novak, M., & Streiner, D. L. (2010). Introduction to health measurement scales. *Journal of Psychosomatic Research, 68*(4), 319–323. https://doi.org/10.1016/j.jpsychores.2010.01.006

Kline, R. B. (2005). *Principles and practice of structural equation modeling*. The Guilford Press.

Knapp, H. (2018). *Intermediate statistics using SPSS*. Sage Publications, Inc.

Koballa, T. R., & Crawley, F. E. (1985). The influence of attitude on science teaching and learning. *Journal of Research in Science Teaching, 85*(3), 222–232. https://doi.org/10.1111/j.1949-8594.1985.tb09615.x

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*, 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

LOMCE. (2013). *Ley Orgánica 8/2013, de 9 de diciembre, para la mejora de la calidad educativa [Organic Law 8/2013, of December 9, 2013, for the improvement of the quality of education]*.

Maltese, A. V., & Tai, R. H. (2011). Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among U.S. students. *Science Education, 95*(5), 877–907. https://doi.org/10.1002/sce.20441

Miller, B. (2021). Developing interest in STEM careers: The need to incorporate STEM in early education. *School Science and Mathematics, 121*, 10–11. https://doi.org/10.1111/ssm.12497

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology, 63*(7), 737–745.

Mundform, D. J., Shaw, D. G., & Tian, L. K. (2005). Minumum sample size recommendations for conducting factor analyses. *International Journal of Testing, 5*(2), 159–168. https://doi.org/10.1207/s15327574ijt0502_4

Navarro, M., Förster, C., González, C., & González-Pose, P. (2016). Attitudes toward science: Measurement and psychometric properties of the test of science-related attitudes for its use in Spanish-speaking classrooms. *International Journal of Science Education, 38*(9), 1459–1482. https://doi.org/10.1080/09500693.2016.1195521

Scalling Procedures Issues and Application. In *BMC Public Health* (Vol. 5, Issue 1). Sage Publications, Inc. https://ejournal.poltektegal.ac.id/index.php/siklus/article/view/298%0Ahttp://repositorio.unan.edu.ni/2986/1/5624.pdf%0A https://doi.org/10.1016/j.jana.2015.10.005%0Ahttps://doi.org/10.1016/j.jana.2015.10.005%0Ahttp://www.biomedcentral.com/1471-2458/12/58%0A

Newman, I., & McNeil, K. (1998). *Conducting survey research in the social sciences*. University Press of America.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. McGraw-Hill.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, and Computers, 32*(3), 396–402. https://doi.org/10.3758/bf03200807

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education, 25*(9), 1049–1079. https://doi.org/10.1080/0950069032000032199

Paños, E., & Ruiz-Gallardo, J. R. (2021). Attitude toward informal science in the early years and development of leisure time in science (LeTiS), a pictographic scale. *Journal of Research in Science Teaching, 58*(5), 689–720. https://doi.org/10.1002/tea.21675

Peters, G.-J.Y. (2014). The alpha and the omega of scale reliability and validity. *The European Health Psychologist, 16*(2), 56–69.

Potvin, P., & Hasni, A. (2014). Interest, motivation and attitude towards science and technology at K-12 levels: A systematic review of 12 years of educational research. *Studies in Science Education, 50*, 85–129. https://doi.org/10.1080/03057267.2014.881626

Roberson, R. B. I., Elliot, T. R., Chang, J. E., & Hill, J. N. (2014). Exploratory factor analysis in rehabilitation psychology: A content analysis. *Rehabilitation Psychology, 59*(4), 429–438. https://doi.org/10.1037/a0037899

Roberts, D. A., & Bybee, R. W. (2014). Scientific literacy, science literacy, and science education. In N. G. Lederman & S. K. Abell (Eds.), *Handbook of Research on Science Education* (Vol. II, pp. 559–572). Routled.

Said, Z., Summers, R., Abd-El-Khalick, F., & Wang, S. (2016). Attitudes toward science among grades 3 through 12 arab students in Qatar: Findings from a cross-sectional national study. *International Journal of Science Education, 38*(4), 621–643. https://doi.org/10.1080/09500693.2016.1156184

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Summers, R., & Abd-El-Khalick, F. (2018). Development and validation of an instrument to assess student attitudes toward science across grades 5 through 10. *Journal of Research in Science Teaching, 55*(2), 172–205. https://doi.org/10.1002/tea.21416

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics (5th ed)*. Pearson Education.

Toma, R. B. (2021a). Problemas de validez y fiabilidad en los cuestionarios ROSE: revisión sistemática de la producción española [Validity and reliability issues in the ROSE Questionnaire: A systematic review of Spanish literature]. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias, 18*(3), 3102.

Toma, R. B. (2021b). Measuring children's perceived cost of school science: Instrument development and psychometric evaluation. *Studies in Educational Evaluation, 70*, 101009. https://doi.org/10.1016/j.stueduc.2021.101009

Toma, R. B. (2024). Elementary school students' interests and attitudes towards biology and physics. *Journal of Biological Education, 58*(4), 972–983. https://doi.org/10.1080/00219266.2022.2147208

Toma, R. B., & Lederman, N. G. (2022). A comprehensive review of instruments measuring attitudes toward science. *Research in Science Education, 52*(2), 567–582. https://doi.org/10.1007/s11165-020-09967-1

Tytler, R. (2014). Attitudes, identity, and aspirations toward science. In N. G. Lederman & S. K. Abell (Eds.), *Handbook of research on science education* (pp. 82–103). Routledge.

Tytler, R., & Osborne, J. (2012). Attitudes and aspirations towards Science. In B. J. Fraser, K. Tobin, & C. J. McRobbie (Eds.), *Second International Handbook of Science Education* (pp. 597–625). Netherlands: Springer.

van den Hurk, A., Meelissen, M., & van Langen, A. (2019). Interventions in education to prevent STEM pipeline leakage. *International Journal of Science Education, 41*(2), 150–164. https://doi.org/10.1080/09500693.2018.1540897

Xu, X., & Lewis, J. E. (2011). Refinement of a chemistry attitude measure for college students. *Journal of Chemical Education, 88*(5), 561–568. https://doi.org/10.1021/ed900071q

Yánez-Pérez, I., Toma, R. B., & Meneses, J. Á. (2024). The IndagApp mobile app: an inquiry-based science teaching resource. Usability evaluation with pre-service teachers. *Journal of New Approaches in Educational Research, 13*(1), 7. https://doi.org/10.1007/s44322-024-00008-7

Zhang, D., & Campbell, T. (2011). The psychometric evaluation of a three-dimension elementary science attitude survey. *Journal of Science Teacher Education, 22*, 595–612. https://doi.org/10.1007/s10972-010-9202-3