Studies in Educational Evaluation 70 (2021) 101009

Contents lists available at ScienceDirect

Studies in Educational Evaluation

journal homepage: www.elsevier.com/locate/stueduc

Measuring children's perceived cost of school science: Instrument development and psychometric evaluation

Radu Bogdan Toma

Department of Specific Didactics, University of Burgos, c/Villadiego, s/n, Burgos, 09001, Spain

ARTICLE INFO

Keywords:
Confirmatory factor analysis
Exploratory factor analysis
Expectancy-value
Task effort
Loss of valued alternatives

ABSTRACT

This investigation addresses the need for valid and reliable instruments that contribute to understanding the factors that lead to the rejection of science-related studies. We discuss the theoretical and methodological limitations of published attitudes toward science questionnaires and describe the development and validation of a short instrument rooted in the cost construct of the expectancy-value model of achievement motivation. We collected data from a sample of six hundred thirty-two 5th and 6th (M_{age} = 10.87; SD = .76) elementary students in Spain. Exploratory and confirmatory factor analysis revealed a parsimonious structure measuring loss of valued alternatives and task effort cost. Further psychometric evaluation displayed evidence for convergent, discriminant, and concurrent validity. Likewise, the reliability was acceptable for both three-item scales. These findings support the proposed instrument to measure barriers experienced by Spanish children when studying school science.

1. Introduction

Promoting students' interest in science is a major goal worldwide (Tytler & Osborne, 2012). Young children show a great interest in school science. Yet, they end up becoming disinterested at the end of elementary education (Summers & Abd-El-Khalick, 2018; Tytler & Osborne, 2012; Wang & Berlin, 2010). Students enter the science pipeline in elementary school but most end up *leaking out* of it later on (Cannady et al., 2014). This is the case in countries where science-related subjects are not always mandatory (European Commission, 2015; LOMCE, 2013). For example, in Spain, up to 30 % of students will reject science at secondary education (Ardura & Pérez-Bitrián, 2018).

Against this background, the literature signals attitudinal variables influencing students' disinterest in science. Lack of enjoyment, low levels of self-efficacy, or perceiving science as irrelevant are detrimental (Andersen & Ward, 2014; Chachashvili-Bolotin et al., 2016; Palmer et al., 2017; Sellami et al., 2017). Yet, the usefulness of such findings is at stake due to a lack of conceptual clarity and valid measurement tools (Toma & Lederman, 2020; Blalock et al., 2008).

This investigation aims at supplying the shortage of instruments grounded in sound frameworks. It advances a questionnaire rooted in Eccles et al.'s (1983) cost construct, which has a significant role in promoting pathways towards STEM. For example, perceived cost predicted negative attitudes and intentions to drop out of STEM disciplines (Ball, Huang, Cotten et al., 2017; Perez et al. 2014). While perceived cost has attracted international attention recently, to the best of the author's knowledge a cost-related measure is yet to be developed for Spanish-speaking children. Hence, the proposed instruments may advance our understanding of what drives students to drop out of science.

2. Literature review: measuring attitudes toward science

Concerns about attitudes toward science instruments date back decades and persist (Blalock et al., 2008; Munby, 1983; Toma & Lederman, 2020). Criticism includes a lack of conceptual analysis of the included items. Likewise, validity and reliability evidence is scarce. Table 1 reports the findings of a non-exhaustive analysis of recently published instruments.

Attitudinal instruments lack a guiding framework and include a broad spectrum of aspects. There are scarce commonalities among such instruments, which calls into question their usefulness. In the absence of common grounds on defining attitudes toward science, test-developers should provide a rationale for the inclusion of such contrasting dimensions as the desire to become a scientist, family encouragement in science, or value of science (Hillman et al., 2016; Lamb et al., 2012; Sabah et al., 2013).

Besides the lack of a guiding framework, attitudinal instruments lack

E-mail address: rbtoma@ubu.es.

https://doi.org/10.1016/j.stueduc.2021.101009
Received 29 January 2020; Received in revised form 7 March 2021; Accepted 10 March 2021
Available online 16 March 2021

Table 1
Synthesis of recently published attitude toward science instruments.

[a]Test developers	Target population	Items	Subscales/constructs	[b]Theory	Validity	Reliability
Abd-el-Khalick et al. (2015)	Elementary Middle Secondary	32-Likert	1. Attitudes toward science and school science 2. Unfavourable outlook on science 3. Control beliefs about ability in science 4. Behavioural beliefs about the consequences of engaging with science 5. Intentions to pursue science	Yes	Content Construct	CFA reliability = .61–.87
Bennett & Hogarth (2009)	Middle Secondary	25-Likert	1. Disposition towards school science 2. Disposition towards science outside of school	No	Content	No
Guzey et al. (2014)	Elementary	28-Likert	1. Personal and social implications of STEM 2. Learning of science and engineering and the relationship to STEM 3. Learning of mathematics and the relationship to STEM 4. Learning and use of technology	No	Construct	α = .77–.87
Hillman et al. (2016)	Elementary Middle Secondary	40-Likert	1. Attitude towards the subject of science 2. Desire to become a scientist 3. Value of science to society 4. Perception of scientists	No	Content	α = .54–.87
Kennedy et al. (2016)	Middle	7-Likert 3-Semantic differential	1. Enjoyableness of school science 2. Self-efficacy in school science 3. Difficulty of school science 4. Usefulness of school science for career 5. Relevance of school science for every-day life 6. Intent to enrol in further school science	No	Construct	α = .82–.98
Lamb et al. (2012)	Elementary Middle Secondary	21-Likert	1. Family encouragement 2. Peer attitudes toward science 3. Teacher influence 4. Informal learning experiences 5. science classroom experiences	No	Construct	α = .50–.72
Sabah et al. (2013)	Middle	11-Likert	1. Positive affect toward science 2. Self-confidence in learning science 3. Students' valuing science	No	Discriminant Construct	Item reliability = .89
Summers & Abd-El-Khalick (2018)	Elementary Middle Secondary	30-Likert	1. Attitudes toward science and school science 2. Behavioural beliefs about science 3. Intentions to engage in science 4. Normative beliefs 5. Control beliefs	Yes	Content Construct	CFA reliability = .70–.91
Toma and Meneses-Villagrá (2019a)	Elementary	7-Likert 3-Semantic differential	1. Enjoyableness of school science 2. Self-efficacy in school science 3. Difficulty of school science 4. Usefulness of school science for career 5. Relevance of school science for every-day life 6. Intent to enrol in further school science	No	Content Predictive Concurrent Construct	α = .70; test-retest = .87 Item-total r = .24–.56
Tyler-Wood, Knezek, & Christensen (2010)	Middle	12-Likert	1. Perception of supportive environment for pursuing science career 2. Interest in pursuing education opportunities that lead to science career 3. Perceived importance of a career in science	No	Content Criterion Construct	α = .94
Tyler-Wood et al. (2010)	Middle Undergraduates In-service teachers	25-Semantic differential	1. Perception of science 2. Perception of technology 3. Perception of engineering 4. Perception of mathematics 5. STEM career interest	No	Content Discriminant Construct	α = .84–.93
Wang & Berlin (2010)	Elementary	30-Likert	1. Attitudes towards Science Class	No	Content Construct	α = .93
Zhang & Campbell (2011)	Elementary	28-Likert	1. Student affective feeling about science 2. Student science learning behaviours 3. Student cognitive judgment of science based on their values and beliefs about science	Yes	Construct Concurrent	α = .88 Test-retest = .91

(*continued on next page*)

validity evidence. For example, the absence of content validity questions the relevance of the items as a reflection of attitudes. Likewise, no discriminant validity evidence raises questions about the dimensionality of the measures. Besides, Cronbach's indices are below the minimum cutoff (α > .70) in many instruments, thus, their reliability is at stake (e.g., Abd-el-Khalick et al., 2015; Vázquez & Manassero, 2009). Such limitations reinforce the need for new instruments grounded on parsimonious theoretical frameworks.

3. The cost construct: a promising guiding theoretical framework

Eccles et al.'s (1983) theory provide a comprehensive framework for understanding achievement motivations and career choices. The first key component is named expectancies of success and incudes performance related beliefs about an upcoming task. The second components embodies different task-values: (i) intrinsic/interest value as the anticipated enjoyment from doing the task; (ii) utility value as if the task fits into individuals' present or future plans; (iii) attainment value as to whether the task is perceived to be important for own identity; and (iv) perceived cost, which refers to what needs to be given up and the anticipated effort for task completion (Eccles & Wigfield, 2020, pp. 3–5).

Eccles et al. (1983) operationalized cost was as a task-value dimension. It comprised perceived effort (i.e., effort needed to be successful at a given task), loss of valued alternatives (i.e., valued activities being missed due to engaging in a given task), and psychological cost of failure (i.e., anxiety for potential failure). Yet, recent investigations suggest cost to be a separated dimension from task-values to avoid potential problems of combining constructs with positive and negative valence (Jiang et al., 2018). In this regard, several authors reported partial cost scales, such as the perceived emotional cost measures of Ball, Huang, Rikard et al. (2017) and Luttrell et al. (2010), or the general cost scales of Chiang et al. (2011) and Kosovich et al. (2015). To tap the multidimensionality of the cost construct, Perez et al. (2014) developed a comprehensive instrument that includes effort-related cost, loss of valued alternatives, and psychological cost. Similarly, Flake et al. (2015) further extended the cost construct to four underlying dimensions: task effort, outside effort, loss of valued alternatives, and emotional cost.

This investigation draws on such studies defining cost as a unique trait separated from task values. Building on Flake et al. (2015), we define cost in science education as *the perceived adverse factors or barriers that induce a negative appraisal of school science*[1] . It includes internal beliefs about the salient negative aspects related to school science, regardless of the ability to be successful in such discipline. High achieving students may also perceive school science to require too much effort and loss of valued alternatives. In this sense, given that career aspirations are more malleable during the early stages of the educational system (Caspi et al., 2019; Ellis et al., 2016), it seems vital to reduce the perceived cost of school science during elementary school grades.

4. Method

4.1. Participants

A total of six hundred thirty-two students enrolled in fifth (36.7 %) and sixth (59.8 %) elementary grades in the Spanish educational system were recruited from 36 state-funded schools located in the province of Burgos. Almost half of the sample were girls (48.4 %), and the mean age of the participants was 10.87 years old (*SD* = .76; range 9–12). School grade and age data were missing for 22 participants.

Compulsory education in Spain comprises six grades of elementary education (from 6 to 12 years old), followed by four grades of secondary education (from 12 to 16 years old). Subsequently, students choose whether to pursue a STEM or non-STEM related bachelor's degree. Such studies comprises two additional grades that are preparatory for university admission (from 16 to 18 years old).

The Spanish educational system includes compulsory science subjects in the elementary stage (i.e., Natural Sciences) and the first three years of the secondary stage (i.e., Biology & Geology and Physics & Chemistry). However, science-related subjects are optional during the fourth year of secondary education. Their choice (or not) configures the type of bachelor and university degrees that students are eligible to enroll in. Therefore, students may drop out of the Science pipeline from the age of 14/15 (LOMCE, 2013).

4.2. Instrument development framework

We designed the proposed cost instrument following a three-phase procedure rooted in contemporary recommendations such as the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council for Assessment in Education [AERA, APA, & NCME], 2014).

The first phase consisted of the development of the questionnaire by cross-culturally adapting items from existing cost instruments, which is a procedure used to adapt self-administered questionnaires for their use in a new country, culture, and/or language (e.g., Beaton et al., 2000). The second phase involved content analysis of the initial pool of items to examine content coverage, relevance, and interpretation of the items, which led to a refinement of the questionnaire before large-scale administration. Finally, the last phase comprised the psychometric evaluation of the proposed questionnaire against construct validity and reliability.

During the last phase, construct validity was examined in terms of the hypothesized dimensionality (structural validity), if items of a particular construct are highly correlated to each other (convergent validity), and poorly correlated to items from other constructs (discriminant validity), and whether the focal measure is correlated with other constructs of conceptual convergence (concurrent validity). As for the reliability domain, the degree of interrelatedness among the items of each construct was examined (internal consistency reliability).

Table 1 (*continued*)

[a]Test developers	Target population	Items	Subscales/constructs	[b]Theory	Validity	Reliability
Vázquez & Manassero (2004, 2005, 2009)	Elementary Secondary Preservice	146-Likert	1. Opinions about science and technology 2. Attitudes toward science classes 3. Environmental challenges 4. My future job 5. Out of school experiences	No	[c]Construct	α = .25–.91

[a] Studies were retrieved from the *Web of Science Core Collection* and *Scopus* database using a combination of "attitude* to* science" keywords.
[b] Refers to whether the items were developed according to an existing theory.
[c] Adequate construct validity evidence was not provided for all the 5 constructs.

[1] Appraisal refers to subjective evaluations or judgments (Flake et al., 2015).

4.2.1. Scale construction

We developed an initial pool of 19 items based on Kosovich et al. (2015) and Flake et al. (2015) cost-related instruments, which are consistent with recent conceptualizations of the cost construct. The selected items were translated into Spanish using a cross-cultural adaptation procedure (Beaton et al., 2000). One bilingual professor translated the items from English to Spanish. Another bilingual professor back translated the Spanish items into the original language. Finally, both professors jointly reviewed the equivalence between the original and the back-translated version.

Selected items were originally developed for middle and undergraduates students. Thus, few changes were made. First, since the “outside effort cost” items included in the Flake et al. (2015) instrument relate to other obligations and commitments (i.e. “I have so many other commitments that I can’t put forth the effort needed for this class”), it was decided to exclude this cost dimension to be more conceptually consistent with the population under study (i.e. elementary students whose commitments are decided by their parents and have little to none personal decisions about them). Second, items were slightly modified to address explicitly the perceived cost of school science and not any other class in general (i.e., “This class is too stressful” was modified to “School science classes are too stressful”). Third, few items were positively worded to enhance children comprehension. Finally, in contrast to the 9-point scale response option of Flake et al. (2015), a 5-point response scale (i.e. totally disagree, disagree, not sure, agree, totally agree) was adopted for simplicity and readability reasons. In this sense, Simms et al. (2019) revealed no improvements in psychometric properties of the measurement instruments beyond six response options.

4.2.2. Scale refining

Table 2 reports the initial list of items and the scale refining process. Before large scale administration, the items were assessed against content validity. A panel of experts composed of two university professors (one from the field of psychology, familiar with the EVT theory and one with expertise in science education and elementary school level) and six elementary school teachers assessed the content validity of the scale. Each expert was provided with the initial item pool and was asked to determine the appropriateness (in terms of construct coverage and readability) of every item for measuring elementary students’ negative appraisals related to studying science using a dichotomous scale (Yes - No). Items rated as being appropriate by at least 6 out of the 8 experts, which is the equivalent of 75 % of inter-rater agreement between the experts (Stemler, 2004), were retained. This process led to a sample of 10 items. There were three task effort items, three items related to loss of valued alternative, and four items measuring emotional negative appraisals.

Next, cognitive think-aloud interviewers were performed with the target population to assess item comprehensibility and interpretation (Beatty & Willis, 2007). A total of 16 students enrolled in 3rd to 6th elementary grades were given a copy of the 10 items retained by the panel of experts and were individually prompted to explain what they were thinking when reading and answering each item. This process revealed that students easily understood all items measuring task effort and loss of valued alternatives perceived cost. However, the emotional cost items were problematic, and students consistently showed difficulties in understanding words like exhausting, frustrating, anxious, or stressful. Therefore, it was decided that such items are not appropriate for the target population and consequently were discarded. Taken together, six items remained for the large-scale administration (Table 2). An analysis of the readability of the items using Fernández-Huerta’ s (1959) corrected formula for Spanish texts, which is based on the Flesch-Kincaid readability test, revealed a readability index of 75.66 (somewhat easy), which corresponds to 4.6 years of school requires according to Crawford’ s index (1984). These findings confirms that the Spanish translation of the items are appropriate for fifth and sixth graders.

Table 2
Items excluded and retained after content validity analysis.

Original items	Initial item pool	[b]Expert panel	[c]Target sample
Task effort cost		–	–
This class demands too much of my time	School science classes demands too much of my time	–	–
I have to put too much energy into this class	I must put too much energy into school science classes	–	–
This class takes up too much time	School science classes takes up too much time	–	–
This class is too much work	School science classes is too much work	–	–
This class takes too much effort	School science classes requires too much effort	Retained	Retained
[a]My [math or science] classwork requieres too much time	[a]My science classwork requires too much time	Retained	Retained
[a]I’m unable to put in the time needed to do well in my [math or science] class	[a]I cannot put in time needed to do well in my Science class	Retained	Retained
Loss of valued alternatives cost		–	–
I have to sacrifice too much to be in this class	I must sacrifice a lot of free time to be good at school science classes	Retained	Retained
This class requieres me to give up too many other activities I value	School science classes requires me to give up too many other activities I value	–	–
Taking this class causes me to miss out on other things I care about	Studying school science classes causes me to miss out on too many other things I care about	–	–
I can’t spend as much doing the other things I would like because I am taking this class	I can’t do other things that I would like because I am studying for school science classes	–	–
[a]I’m unable to put in the time needed to do well in my [math or science] class	[a]I must invest a lot of time to get good grades in science	Retained	Retained
[a]I have to give up too much to do well in my [math or science] class	[a]I must give up too much to do well in my science class	Retained	Retained
Emotional cost		–	–
I worry too much about this class	I worry too much about school science classes	Retained	–
This class is mentally exhausting	School science classes are too exhausting	–	–
This class is emotionally draining	School science classes are emotionally draining	–	–
This class is too frustrating	School science classes are too frustrating	Retained	–
This class is too stressful	School science classes are too stressful	Retained	–
This class makes me feel too anxious	School science classes makes me feel too anxious	Retained	–

[a] Items adapted from Kosovich et al. (2015, 22). The rest of the items are adapted from Flake et al. (2015, p. 239).
[b] Refers to retained items after the panel of experts analyzed them.
[c] Refers to retained items after performing think-aloud interviews with the target sample – refers to items that were rejected after panel of expert analysis or cognitive interviews with the target sample.

4.2.3. Psychometric analyses

The structural validity of the proposed instrument was assessed in two stages by randomly splitting the sample into two groups. Responses from the first subgroup ($n = 327$, 48.3 % girls) were subjected to robust exploratory factor analysis using Unweighted Least Squares extraction procedure on Polychoric rather than Pearson correlations matrices, as this procedure provides more accurate results with ordinal-Likert type data (Gaskin & Happell, 2014; Holgado-Tello et al., 2010). Factors were rotate using Promax oblique rotation, as it is preferred over orthogonal rotation (e.g. Varimax) when factors are expected to be correlated (Roberson et al., 2014). The number of extracted factors was determined based on the results of BIC dimensionality test, parallel analysis, and the Minimum Average Partial (MAP) test using the SPSS v.25 (O'Connor, 2000) and FACTOR software (Lorenzo-Seva & Ferrando, 2006).

Next, responses from the second subgroup (n = 305, 48.5 % girls) were subjected to confirmatory factor analysis to examine three different models. The first model was composed of a unique, general factor measuring perceived cost. The second model consisted of two, first-order factors measuring *task effort* and *loss of valued alternatives* costs. The last model consisted of a first-order factor composed of two related sub-dimensions. Given that skewness (Sk = .237–.822) and kurtosis (k = −.902 to −.065) values were smaller than 1, Maximum Likelihood (ML) estimation method was used (Byrne, 2010). Each model was assessed against the following goodness-of-fit criteria (Byrne, 2010): (i) the root mean square error of approximation (RMSEA) of .05 or lower (ii) the goodness of fit index (GFI) of .95 or greater, (iii) the comparative fit index (CFI) of .95 or greater, and (iv) the parsimony goodness of fit index (PGFI) lower than .50. Analyses were performed using AMOS v.23 software (Arbuckle, 2014).

Concurrent validity was established using the total sample (N = 632) by examining the relationship between students' perceived cost and their attitudes towards school science, measured using (Toma & Meneses-Villagrá, 2019a, 2019b) Spanish School Science Attitude Survey (S-SSAS). This instrument consists of six unique, but conceptually related scales comprising different affective, cognitive, and conative aspects of students' attitudes to school science using single-items with: (i) intention to enroll in further science, (ii) enjoyableness, (iii) perceived difficulty, (iv) self-concept, (v) usefulness, and (vi) relevance of school science. For this study, only three attitudinal dimensions of conceptual convergence with expectancies of success (i.e., self-concept and perceived difficulty of school science) and intrinsic/interest values (i.e. enjoyableness of school science) were administered.

Specifically, self-concept was measured through the item "I think I am very good at science" (1 – Strongly disagree; 5 – Strongly agree); the perceived difficulty of school science was examined using the item "I struggle with completing the assignments for science class" (1 – Strongly disagree; 5 – Strongly agree); and enjoyableness of school science was assessed through the item "I think science is (1 – Boring; 5 – Fun)" (Toma & Meneses-Villagrá, 2019a, p. 8). Consistent with the EVT literature that considers perceived cost to be negatively related to students expectancies of success and task values (for two exhaustive reviews, see Eccles & Wigfield, 2020; Wigfield & Eccles, 2020), the proposed cost instrument is expected to be negatively correlated to students' enjoyableness of school science and their self-concept in school science, and positively correlated with their perceived difficulty of school science.

Finally, using the total sample (N = 632) the internal consistency reliability of the proposed cost instrument was established against (i) Cronbach alpha (α) of .70 or greater, (ii) Spearman-Brown split-half reliability coefficient above .60, (iii) greatest lower bound to reliability (glb) above .70, and (iv) McDonald's omega (ω) of .70 or greater.

5. Results

5.1. Exploratory factor analysis

The Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) was .79 (fair) and Bartlett's Test of Sphericity was statistically significant, thus supporting the factorability of the data. While the MAP test suggested that only one factor should be extracted, the BIC dimensionality test and results of parallel analysis (Fig. 1) suggested that two factors exceed the corresponding criterion values of a randomly generated data matrix. The decision to extract two factors was further supported by the parsimonious conceptual distinction between both factors that were consistent with the hypothesized structure.

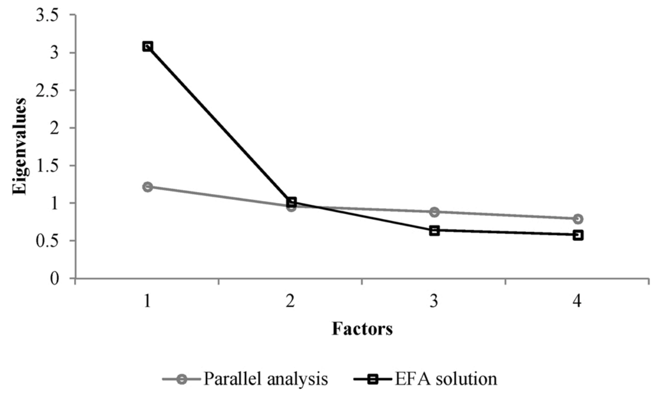

Fig. 1. Parallel analysis results for factor extraction decision.

Using a polychoric (tetrachoric) correlation matrix, the robust factor analysis with the Unweighted Least Squares (ELS) extraction method revealed two factors with initial eigenvalues above one explaining a total of 68.4 % of the item variance. Promax rotation revealed a simple structure with items strongly loading only on the hypothesized factor, and with no cross-loadings above the .40 criteria between factors. Therefore, the first latent variable, named *Task effort cost*, measures the effort students feel they must perform to be successful in school science. The second factor, named *Loss of valued alternative cost*, measures the extent to which students perceived that they need to give up other valued activities to study school science (Table 3). Extracted reliability of rotated factors was .94 and .90, respectively, and the h-latent index was .816 for the Loss of valued alternatives and .79 for the Task effort construct, suggesting well defined latent factors that are more likely to be reproducible and stable in further studies. Besides, the simplicity index was .99 (95 % IC = .98–.99), indicating that the simplicity in the loading matrix is near perfect.

5.2. Confirmatory factor analysis

The goodness of fit indices from confirmatory factor analysis revealed that the model extracted from exploratory factor analysis fit best the data (Table 4), suggesting that Task effort and Loss of valued alternatives are two conceptually distinct constructs measuring the perceived cost of school science. Pearson correlation coefficient revealed medium to high positive correlations within the items included in the Task-effort construct (ranges from .33 to .55) and within the Loss of valued alternatives items (ranges from .42 to .49), which provides empirical support for convergent validity. Likewise, Pearson correlation coefficient for the complete dataset (r = .51) and both EFA (r = .49) and CFA (r = .66) inter-factor correlation matrix for the data sub-sets revealed that both factors are correlated, however, since correlations between factors were lower than cut-off criteria of r < .80 (Brown, 2006), the discriminant validity between Task-effort and Loss of valued alternatives constructs is empirically confirmed. In this sense, it should

Table 3
Polychoric (tetrachoric) pattern matrix.

Items	Factors 1	2	h^2
1. My Science classwork requires too much time	**.76**	.11	.67
2. I cannot put in the time needed to do well in my Science class	**.88**	.20	.64
3. I must give up too much to do well in my science class	.37	**.53**	.61
4. I must invest a lot of time to get good grades in science	.10	**.79**	.72
5. Science class requires too much effort	**.80**	.06	.69
6. I must sacrifice a lot of free time to be good at science	.21	**.97**	.78

h^2 = item communalities. Factor 1 = Task effort; Factor 2 = Loss of valued alternatives.

Table 4
Goodness of fit statistics.

Model	χ^2	*p*	RMSEA	GFI	CFI	PGFI
One-factor model	78.33	< .01	.16	.91	.85	.39
Two-factor model	8.56	.29	.03	.99	.99	.33
Second order two factor model	11.14	.13	.04	.99	.99	.32

χ2 = chi-square; RMSEA = Root mean square error of approximation; GFI = Goodness of fit index; CFI = Comparative fit index; PGFI = Parsimony goodness of fit index.

be noted that the model composed of a second-order factor with two distinct dimensions (Fig. 2) displayed also adequate fit indices, suggesting that a single cost score can be calculated by summing the scores of the Task effort and Loss of valued alternatives scales.

5.3. Concurrent validity

Pearson correlation coefficient revealed that both cost constructs were significantly related to the attitudes toward science dimensions with which conceptual convergence was expected. More specifically, there was a negative relationship between perceived cost and enjoyableness of school science and self-concept in school science, and a positive relationship between perceived cost and perceived difficulty of school science (Table 5). These findings provide evidence of concurrent validity for the proposed instrument. It should also be noted that all response options have been selected for each item, and that the mean of this population is in the middle of the scale (2.33–2.44 out of 5),

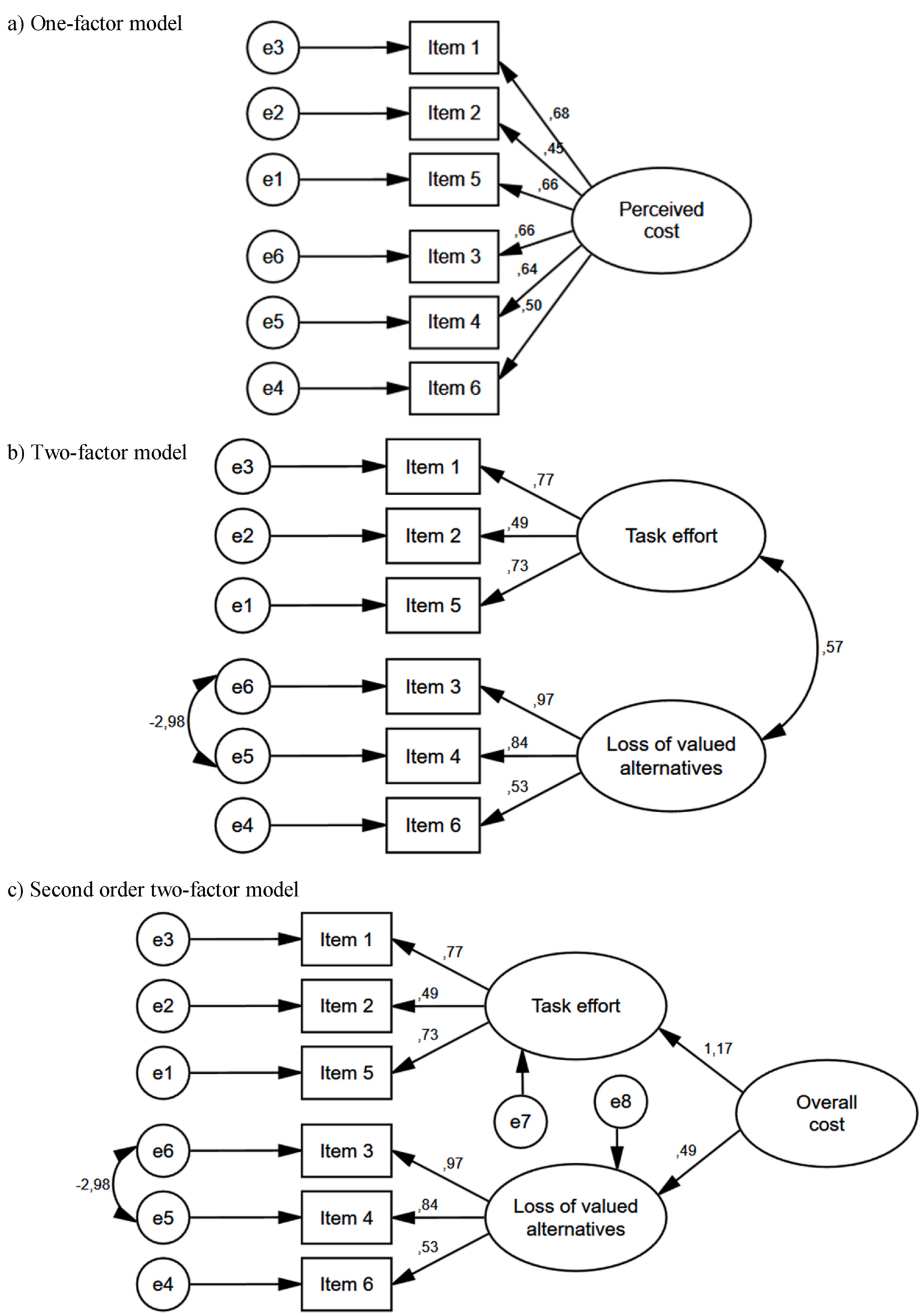

Fig. 2. Competing factor models with standardized coefficients.

Table 5
Pearson correlation between perceived cost and attitudes toward science.

Cost measures	*M*	*SD*	External measures		
			Enjoyableness	Self-concept	Difficulty
Task effort	2.33	.87	−.33	−.33	.27
Loss of valued alternatives	2.55	.98	−.14	−.26	.22
[a]Overall cost	2.44	.80	−.27	−.34	.28

All correlations were significant .01 level.
[a] Refers to combined scores of Task effort and Loss of valued alternatives constructs.

suggesting that there is no floor or ceiling effect.

5.4. Reliability

The internal consistency of the Task effort and Loss of valued alternatives constructs was acceptable, as indicated by Cronbach α of .78 and .70, and Spearman-Brown coefficient of .74 and .72, respectively. When computing a single overall score, the reliability of the proposed instrument improves, as indicated by Cronbach α of .81, McDonald's ω of .82, and the greatest lower bound to reliability (glb) of .86.

6. Discussion

This study intended to fill the gap in conceptually robust and psychometrically valid and reliable measurement instruments that could be helpful in understanding variables affecting elementary school students' interest in science. It is argued that recent conceptualizations of the perceived cost construct first introduced in Eccles et al.' s (1983) expectancy-value theory could be used to understand the factors that affect students' intention to enroll in science-related optional studies and pursue STEM-related university degrees. Since middle and secondary school represents the start of the leakages in the Science pipeline (Ball, Huang, Cotten et al., 2017 Ball, Huang, Rikard et al., 2017; Toma & Meneses-Villagrá, 2019b), the current investigation presents the development and validation of a short instrument that can be used to examine the perceived cost of school science at the end of elementary grades.

This instrument intends to overcome some of the conceptual and methodological limitations found in attitudes toward science instruments, such as lack of a guiding theoretical framework for item development, the use of a broad spectrum of dimensions that are not conceptually justified, or the underreporting of validity and reliability evidence. Using a three-phase procedure for item selection, scale refining, and psychometric evaluations, the proposed instrument showed promising evidence of content and construct validity (i.e. structural, convergent, discriminant, and concurrent), as well as satisfactory internal consistency reliability.

When compared to the existing attitudes toward science instruments, the proposed instrument has several advantages and strengths. The procedure followed to develop and validate the proposed instruments attempted to overcome the criticisms raised in the science education research literature concerning existing instruments (Blalock et al., 2008; Toma & Lederman, 2020).

Specifically, the items were rooted in a comprehensive theoretical framework. Next, such items were assessed against several psychometric indices. In this sense, the analyses used to examine the structural validity adhered to contemporary recommendations that consist of first exploring the underlying factor structure of the retained items through EFA, and subsequently confirming the hypothesized structure by analyzing whether the model fits the data properly through CFA (Lloret-Segura et al., 2014). Besides, although Pearson correlation matrices are widely used for factor analysis of Likert-type items, a Polychoric correlation matrix was used instead since it is best suited for the ordinal type of items (Holgado-Tello et al., 2010).

Likewise, the structural validity of most attitudes toward science instruments was determined using the strongly discouraged Little Jiffy procedure (consisting of principal component extraction with orthogonal Varimax rotation and factor retention based on Kaiser criterion > 1), which has been criticized in the literature for the lack of robustness in the results it provides and for extracting more dimensions than those that underlie the construct under study (Gaskin & Happell, 2014; Lloret-Segura et al., 2014). Nonetheless, the procedure used in this study to determine the structural validity of the cost instrument followed current recommendations that advocate for robust factorial extraction methods, oblique rotations, and the use of several criteria for factor retention decision (Gaskin & Happell, 2014). Hence, the cost questionnaire reported in this study can be postulated as an alternative to the conceptually confusing attitudes towards science construct for the investigation of those factors affecting students' choice of scientific studies and careers.

6.1. Validity and reliability evidence

The use of a panel of experts and think-aloud interviews with the target population indicated that the items included in the proposed instrument were easily understood and interpreted by children enrolled in elementary education, thus providing support for content validity. Next, exploratory, and confirmatory factor analysis provided empirical evidence of structural validity and supported the existence of a simple structure composed of two well-defined and parsimonious underlying factors measuring task effort and loss of valued alternatives costs. These findings are consistent with recent studies that conceptualize cost as a multidimensional factor separate from the task-value component included in the expectancy-value model of achievement motivation (i.e. Flake et al., 2015; Perez et al., 2014).

Both perceived cost construct displayed evidence of convergent and discriminant validity, thus they can be used as two conceptually distinct factors that measure different types of the perceived cost. Likewise, the results of this study suggest that scores of both factors can be added to obtain an overall cost measure. Finally, consistent with expectancy-value literature, both cost factors were positively related to perceived difficulty, and negatively related to enjoyableness and self-concept in school science, thus providing initial evidence for concurrent validity.

Reliability indices reached satisfactory results, with internal consistency results above the minimum cut-off recommended for preliminary research and consistent with recommendations related to scales with limited items or aimed for young respondents (Newman & McNeil, 1998; Nunnally & Bernstein, 1994). Therefore, the proposed instrument can be considered as a promising first step into the measurement of the perceived cost of school at the elementary school level.

6.2. Limitations and avenues for future research

Three main limitations must be acknowledged. First, the proposed cost instrument is composed of items originally developed for middle and undergraduates students (Flake et al., 2015; Kosovich et al., 2015). This decision proved to be problematic, since the items measuring emotional cost were not correctly interpreted by the elementary school students included in this study, thus reducing the factor structure of the proposed instrument to two types of perceived cost: task effort and loss of valued alternatives.

Second, while the two factor and the second-order two-factor models displayed great fit indices, covariance between the errors of two items in the "loss of valued alternatives" subscale were necessary, which may indicate that these items are very similar (both refer to being successful in school science). Finally, concurrent validity results should be interpreted considering that the external measures are composed of single-item scales, which are less stable, reliable, and accurate than multipleitem scales (Bowling, 2005). Therefore, further studies examining the concurrent validity of the cost measure proposed using valid and reliable expectancies of success and task-value instruments are encouraged. Given that such measures are not available for Spanish speaking children, the concurrent results reported in this study represent a first and promising step towards such an endeavor.

Despite these limitations, this study has several implications for future research. Given its simple structure and short administration time, the proposed instrument can be used to examine students' perceived cost of school science during the last years of elementary education. Therefore, the instrument advanced in this study paves the way for future research on how to assess a valuable variable for repairing the leaky pipeline for the early stages of the educational system, thus being a promising questionnaire for understanding students' (des)interest in science-related careers. Future studies examining the psychometric properties of the proposed instrument on high school students are strongly encouraged, which would provide a measurement tool that can be used in longitudinal studies assessing the development and evolution of costs related to school science during compulsory education. Finally, a promising line of research is to extend this cost instrument by including items related to emotional costs experienced by students. Due to the young nature of the sample used in this study, the emotional items adapted from Flake et al. (2015) failed at being easily understood by the children, thus being excluded from the final version. Therefore, future studies are warranted that examine the possibility of developing an emotional cost scale related to school science, operationalized so that young students can understand it.

Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgments

I am grateful to the anonymous referees for their constructive comments.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.stueduc.2021.101009.

References

Abd-el-Khalick, F., Summers, R., Said, Z., Wang, S., & Culbertson, M. (2015). Development and large-scale validation of an instrument to assess Arabic-speaking students' attitudes toward science. *International Journal of Science Education, 37*(16), 2637–2663. https://doi.org/10.1080/09500693.2015.1098789.

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME) (Eds.). (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Andersen, L., & Ward, T. J. (2014). Expectancy-value models for the STEM persistence plans of ninth-grade, high-ability students: A comparison between black, hispanic, and white students. *Science Education, 98*(2), 216–242. https://doi.org/10.1002/sce.21092.

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

Ardura, D., & Pérez-Bitrián, A. (2018). The effect of motivation on the choice of chemistry in secondary schools: Adaptation and validation of the Science Motivation Questionnaire II to Spanish students. *Chemistry Education Research and Practice, 19*(3), 905–918. https://doi.org/10.1039/c8rp00098k.

Ball, C., Huang, K. T., Cotten, S. R., & Rikard, R. V. (2017). Pressurizing the STEM pipeline: An Expectancy-Value theory analysis of youths' STEM attitudes. *Journal of Science Education and Technology, 26*(4), 372–382. https://doi.org/10.1007/s10956-017-9685-1.

Ball, C., Huang, K. T., Rikard, R. V., & Cotten, S. R. (2017). The emotional costs of computers: An expectancy-value theory analysis of predominantly low-socioeconomic status minority students' STEM attitudes. *Information, Communication and Society, 0*(0), 1–24. https://doi.org/10.1080/1369118X.2017.1355403.

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine, 25*(24), 3186–3191. https://doi.org/10.1097/00007632-200012150-00014.

Beatty, P. C., & Willis, G. B. (2007). Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly, 71*(2), 287–311. https://doi.org/10.1093/poq/nfm006.

Bennett, J., & Hogarth, S. (2009). Would you want to talk to a scientist at a party? High school students' attitudes to school science and to science. *International Journal of Science Education, 31*(14), 1975–1998. https://doi.org/10.1080/09500690802425581.

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.

Bowling, A. (2005). Just one question: If one question works, why ask several? *Journal of Epidemiology and Community Health, 59*(5), 342–345. https://doi.org/10.1136/jech.2004.021204.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (Vol. 2). Routledge Taylor & Francis Group. https://doi.org/10.4324/9781410600219.

Cannady, M. A., Greenwald, E., & Harris, K. N. (2014). Problematizing the STEM pipline metaphor: Is the STEM pipeline metaphor serving our students and the STEM workforce? *Science Education, 98*(3), 443–460. https://doi.org/10.1002/sce.21108.

Caspi, A., Gorsky, P., Nitzani-Hendel, R., Zacharia, Z., Rosenfeld, S., Berman, S., & Shildhouse, B. (2019). Ninth-grade students' perceptions of the factors that led them to major in high school science, technology, engineering, and mathematics disciplines. *Science Education, 103*(5), 1176–1205. https://doi.org/10.1002/sce.21524.

Chachashvili-Bolotin, S., Milner-Bolotin, M., & Lissitsa, S. (2016). Examination of factors predicting secondary students' interest in tertiary STEM education. *International Journal of Science Education, 38*(3), 366–390. https://doi.org/10.1080/09500693.2016.1143137.

Chiang, E. S., Byrd, S. P., & Molin, A. J. (2011). Children's perceived cost for exercise: Application of an expectancy-value paradigm. *Health Education &Behavior: The Official Publication of the Society for Public Health Education, 38*(2), 143–149. https://doi.org/10.1177/1090198110376350.

Crawford, A. N. (1984). *A Spanish-language fry-type readability procedure: Elementary level*. Bilingual Education Paper Series. California State University, Los Angeles: Evaluation, Dissemination, and Assessment Center.

Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology, 61*(May), Article 101859. https://doi.org/10.1016/j.cedpsych.2020.101859.

Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values and academic behaviours. In J. T. Spencer (Ed.), *Achievement and achievement motivation* (pp. 75–146). W. H. Friedman.

Ellis, J., Fosdich, B. K., & Rasmussen, C. (2016). Women 1.5 times more likely to leave STEM pipeline after calculus compared to men: Lack of matehamtical confidence a potential culprit. *PloS One, 11*(7), Article e0157447. https://doi.org/journal.pone.0157447.

European Commission. (2015). *Science education for responsible citizenship*. Directorate-General for Research and Innovation. https://doi.org/10.2777/12626.

Fernández-Huerta, J. (1959). Medidas sencillas de lecturabilidad [Simple measures of readability]. *Consigna, 214*, 29–32.

Flake, J. K., Barron, K. E., Hulleman, C., McCoach, B. D., & Welsh, M. E. (2015). Measuring cost: The forgotten component of expectancy-value theory. *Contemporary Educational Psychology, 41*, 232–244. https://doi.org/10.1016/j.cedpsych.2015.03.002.

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.

Guzey, S. S., Harwell, M., & Moore, T. (2014). Development of an instrument to assess attitudes toward Science, Technology, Engineering, and Mathematics (STEM). *School Science and Mathematics, 114*(6), 271–279. https://doi.org/10.1111/ssm.12077.

Hillman, S. J., Zeeman, S. I., Tilburg, C. E., & List, H. E. (2016). My Attitudes toward Science (MATS): The development of a multidimensional instrument measuring students' science attitudes. *Learning Environments Research, 19*, 203–219. https://doi.org/10.1007/s10984-016-9205-x.

Holgado-Tello, F. P., Chacón-Moscoso, S., Barbero-García, I., & Vila-Abad, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality and Quantity, 44*(1), 153–166. https://doi.org/10.1007/s11135-008-9190-y.

Jiang, Y., Rosenzweig, E. Q., & Gaspard, H. (2018). An expectancy-value-cost approach in predicting adolescent students' academic motivation and achievement. *Contemporary Educational Psychology, 54*(June), 139–152. https://doi.org/10.1016/j.cedpsych.2018.06.005.

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.
Kosovich, J. J., Hulleman, C. S., Barron, K. E., & Getty, S. (2015). A practical measure of student motivation: Establishing validity evidence for the expectancy-value-cost scale in middle school. *The Journal of Early Adolescence, 35*(5–6), 790–816. https://doi.org/10.1177/0272431614556890.
Lamb, R. L., Annetta, L., Meldrum, J., & Vallett, D. (2012). Measuring science interest: Rasch validation of the science interest survey. *International Journal of Science and Mathematics Education, 10*, 643–668. https://doi.org/10.1007/s10763-011-9314-z.
Lloret-Segura, S., Ferreres-Traves, A., Hernández-Baeza, A., & Tomás-Marco, I. (2014). El análisis factorial exploratorio de los ítems: una guía práctica, revisada y actualizada [Exploratory item factor analysis: A practical guide revised and updated]. *Anales de Psicología, 30*(3), 1151–1169. https://doi.org/10.6018/analesps.30.3.199361.
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]*.
Lorenzo-Seva, U., & Ferrando, P. J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavior Research Methods, 38*(1), 88–91. https://doi.org/10.3758/BF03192753.
Luttrell, V. R., Callen, B. W., Allen, C. S., Wood, M. D., Deeds, D. G., & Richard, D. C. S. (2010). The mathematics value inventory for general education students: Development and initial validation. *Educational and Psychological Measurement, 70*(1), 142–160. https://doi.org/10.1177/0013164409344526.
Munby, H. (1983). Thirty studies involving the "Scientific Attitude Inventory": What confidence can we have in this instrument? *Journal of Research in Science Teaching, 20*(2), 141–162.
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 & Computers, 32*(3), 396–402. https://doi.org/10.3758/bf03200807.
Palmer, T.-A., Burke, P. F., & Aubusson, P. (2017). Why school students choose and reject science: A study of the factors that students consider when selecting subjects. *International Journal of Science Education, 39*(6), 645–662. https://doi.org/10.1080/09500693.2017.1299949.
Perez, T., Cromley, J. G., & Kaplan, A. (2014). The role of identity development, values, and costs in college STEM retention. *Journal of Educational Psychology, 106*(1), 315–329. https://doi.org/10.1037/a0034027.
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.
Sabah, S., Hammouri, H., & Akour, M. (2013). Validation of a scale of attitudes toward science across countries using rasch model: Findings from TIMSS. *Journal of Baltic Science Education, 12*(5), 692–703.
Sellami, A., El-Kassem, R. C., Al-Qassass, H. B., & Al-Rakeb, N. A. (2017). A path analysis of student interest in STEM, with specific reference to Qatari students. *Eurasia Journal of Mathematics, Science and Technology Education, 13*(9), 6045–6067. https://doi.org/10.12973/eurasia.2017.00999a.
Simms, L. J., Zelazny, K., Williams, T. F., & Bernstein, L. (2019). Does the number of response options matter? Psychometric perspectives using personality questionnaire data. *Psychological Assessment, 31*(4), 557–566. https://doi.org/10.1037/pas0000648.
Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Education, 9*(4), 1–11.
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.
Toma, R. B., & Lederman, N. G. (2020). A comprehensive review of instruments measuring attitudes toward science. *Research in Science Education*. https://doi.org/10.1007/s11165-020-09967-1.
Toma, R. B., & Meneses-Villagrá, J. A. (2019a). Validation of the single-items Spanish-School Science Attitude Survey (S-SSAS) for elementary education. *PLoS One, 14*(1), e0209027. https://doi.org/10.1371/journal.pone.0209027.
Toma, R. B., & Meneses-Villagrá, J. A. (2019b). Preferencia por contenidos científicos de física o de biología en Educación Primaria: un análisis clúster [Preference for physics or biology scientific content in Elementary Education: a cluster analysis]. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias, 16*(1), 1104. https://doi.org/10.25267/Rev_Eureka_ensen_divulg_cienc.2019.v16.i1.1104.
Tyler-Wood, T., Knezek, G., & Christensen, R. (2010). Instruments for assessing interest in STEM content and careers. *Journal of Technology and Teacher Education, 18*(2), 341–363. https://doi.org/10.1016/j.tate.2005.03.006.
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. https://doi.org/10.1007/978-1-4020-9041-7.
Vázquez, A., & Manassero, M. A. (2004). Imagen de la ciencia y la tecnología al final de la educación obligatoria [Image of science and technology at the end of compulsory education]. *Cultura y Educación, 16*(4), 385–398.
Vázquez, A., & Manassero, M. A. (2005). La ciencia escolar vista por los estudiantes [School science as seen by students]. *Bordón. Revista de Pedagogía, 57*(5), 125–144.
Vázquez, A., & Manassero, M. A. (2009). Factores actitudinales determinantes de la vocación científica y tecnológica en secundaria [Attitudinal determinants of scientific and technological vocation in secondary school]. *Cultura y Educacion, 21*(3), 319–330. https://doi.org/10.1174/113564009789052280.
Wang, T. L., & Berlin, D. (2010). Construction and validation of an instrument to measure Taiwanese elementary students' attitudes toward their science class. *International Journal of Science Education, 32*(18), 2413–2428. https://doi.org/10.1080/09500690903431561.
Wigfield, A., & Eccles, J. S. (2020). 35 years of research on students' subjective task values and motivation: A look back and a look forward. In *Advances in motivation science* (1st ed., Vol. 7). Elsevier Inc.. https://doi.org/10.1016/bs.adms.2019.05.002
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.