Studies in Educational Evaluation 74 (2022) 101180

Contents lists available at ScienceDirect

Studies in Educational Evaluation

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

Measuring science capital, science attitudes, and science experiences in elementary and middle school students

M.G. Jones [a], K. Chesnutt [b], M. Ennes [c], D. Macher [d], M. Paechter [d,*]

[a] Department of STEM Education, NC State University, Raleigh, NC, USA
[b] College Access Partnerships, Appalachian State University, Boone, NC, USA
[c] Department of Natural History, University of Florida, Gainesville, FL, USA
[d] Institute of Psychology, University of Graz, Graz, Austria

ARTICLE INFO

Keywords:
Science capital
Science attitude
Science experiences
Elementary and middle school
Value of science

ABSTRACT

Educators and researchers have been struggling for decades to isolate and measure the factors that contribute to science career decisions. However, there are limited instruments available to measure these factors, particularly across different educational levels. This study describes the adaptation and characteristics of the middle school *NextGen Scientist Survey* as an assessment for elementary students. Internal validity of the survey was investigated using confirmatory factor analysis (CFA). Also, measurement invariance between the surveys for middle school and elementary students and across gender was examined. Confirmatory factor analyses found a 5-factor solution for both surveys. Partial scalar invariance across elementary and middle school students could be confirmed, i.e., the factor means can be compared across groups. Furthermore, full scalar invariance can be assumed for gender. Altogether, the *NextGen Scientist Survey* is a valid assessment that can be used across elementary and middle school educational levels.

1. Introduction

Knowledge and competencies in and attitudes towards science are crucial for young people to be successful in society. The “PISA 2024 Strategic Vision and Direction for Science” regards science knowledge, skills and attitudes as basic requirements to face societal challenges of the future successfully. According to this vision scientific knowledge and competencies reach beyond academic settings and individual needs and have impact on society as a whole, e.g., concerning health, societal prosperity (OECD, 2020). Science education plays a crucial role in fostering such desirable knowledge and competencies and in instilling positive attitudes towards science.

In contrast to the critical role that science plays in society, science is not among the most liked subjects studied in school. Estrangement from science subjects (as well as mathematics) occurs as early as elementary school (Newall et al., 2018). Thus, it is not surprising that students choose other subjects when possible and that many shy away from a career in science (DeWitt, Archer, & Moote, 2019; Kessels, Rau, & Hannover, 2006; Pey-Tee Oon & Subramaniam, 2013). At the same time, in a multitude of countries the low numbers of students choosing to major in science, technology, engineering, or mathematics (STEM) is a growing concern (Ertl, Luttenberger, & Paechter, 2017; Jones, Ennes, Weedfall, Chesnutt, & Cayton, 2021) because the demand for a qualified workforce in STEM fields cannot be met.

Given this context, schools are faced with understanding students’ attitudes towards science and counseling them about possible science careers. For these purposes, assessments are needed that can capture a student’s attitudes towards, experiences, and self-beliefs in science and science careers from an early age on, preferably beginning with elementary school children. However, up to now there is a dearth of suitable standardized measurement instruments. The present study describes the development of an instrument that captures factors that contribute to science career aspirations. Moreover, it can be used for different age groups, specifically children in elementary and in middle schools.

The following sections outline the development of students’ science-related attitudes and beliefs (as a theoretical basis for the development of an instrument) as well as a description of existing related instruments, including their advantages and shortcomings.

* Correspondence to: Karl-Franzens-Universitaet, Institute of Psychology, Universitaetsplatz 2, 8010 Graz, Austria.
E-mail address: manuela.paechter@uni-graz.at (M. Paechter).

https://doi.org/10.1016/j.stueduc.2022.101180
Received 1 July 2021; Received in revised form 9 April 2022; Accepted 16 June 2022
Available online 2 July 2022

1.1. Development of students' science-related attitudes and beliefs

As early as childhood, individuals form beliefs about their abilities and interest in different subjects, topics, and possible future pathways. Children as young as six years old establish preferences for specific academic domains and shy away from others (Van Tuijl & van der Molen, 2015). Gender-differences in preferences for and emotions towards specific subjects like mathematics have also been observed in the early years (Luttenberger et al., 2018).

Children's attitudes towards science are formed by their experiences in educational institutions (from elementary education onward), in the family, and in the wider social context (Nugent et al., 2015). Learning to participate in science is a form of cultural capital developed in childhood through engagement with the parents and in out-of-school science experiences over time (Claussen & Osborne, 2013). In the home environment, parents and other significant adults provide science experiences that may stimulate learning and shape children's attitudes toward science. Parents may also serve as role models and influence children by their own attitudes toward science (Jacobs & Bleeker, 2004; Luttenberger, Wimmer, & Paechter, 2018; Rodriguez-Planas & Nollenberger, 2018; Watt, 2004; Watt, Bucich, & Dacosta, 2019). Factors such as parents' and the wider family's attitudes towards and competencies in science, parents' occupations, their expectations for their children, the availability of networks of adults who work in STEM fields, and available science resources comprise what has been described as a child's science capital (Archer et al., 2012).

Outside of the home, the concept of science capital distinguishes academic, social, and cultural aspects as dimensions of a student's life that may or may not encourage interest and participation in science (Archer et al., 2015). Didactic instructional strategies that foster positive attitudes towards science have the potential to be successful in encouraging students in science tasks (UNESCO, 2017; Wentzel & Wigfield, 2007). It seems to be important that interventions to foster interest and raise career aspirations in science start early, i.e., in elementary and early lower secondary education (Blanchard & Lichtenberg, 2003; Luttenberger et al., 2019).

The concept of science capital provides some explanations for why children from different socio-economic backgrounds tend to have varying levels of success and achievement in school (Claussen & Osborne, 2013). Differences in the socio-economics of families influence the science capital of young people through exposure, or a lack thereof, to career pathways or opportunities for engagement in science practices between parent and child (Archer et al., 2012).

Within the broad concept of science capital, students make choices related to interests, experiences, and careers. Expectancy-value theory is another widely used concept that has been used to explore the motivational factors that influence students' academic choices (Eccles & Wigfield, 2020) including career decisions (e.g., Lent & Brown, 2019).

Expectancy-value theory is based on the premise that the individual's perceived or subjective value is crucial for undertaking an activity. Wigfield and Eccles (2000) distinguish between different types of values: intrinsic value, utility value, and attainment value. Intrinsic value is related to an individual's satisfaction in doing the activity; utility value includes the degree to which a person sees the activity as related to goals, while attainment value includes the perception of value of the activity (such as prestige). Furthermore, perceived costs are important factors that influence whether an individual will undertake an activity. They include the effort required to pursue an activity, time, and other variables such as emotional investment (Eccles & Wigfield, 2020; Wigfield & Eccles, 2000).

Experiences also play a major role within expectancy-value theory (Wigfield & Eccles, 2000). In the case of science, this would mean positive in- and out-of-school experiences may raise favorable attitudes towards, interest in, and career aspirations for science. Such experiences can take different forms and may take place in different situations and settings, from family activities to school. Research related to science capital (Archer et al., 2012) as well as studies of science hobbyists (Jones et al., 2019) have shown that knowing someone who works in a STEM field, or does STEM in their leisure time, can shape students' interests and engagement in science (Jones et al., 2021). If a child does not know anyone who does science in a job, it is less likely that that child will develop an early interest in science and attach a value to science. Positive experiences like success in solving tasks also contribute to a positive self-concept (the self-assessment in a particular domain in comparison to a frame of reference, e.g., co-learners, one's former achievement, or a performance criterion [Wimmer et al., 2020]) and high self-efficacy (the belief what one can accomplish in a particular situation (Marsh et al., 2019).

1.2. Need for instrumentation for assessing students' science-related attitudes and beliefs

There are various barriers for young people on the path to a career in science. They may lie within the family, in the wider social context, and/or in the person, e.g., as an overly critical concept of one's own abilities or a pessimistic view about one's career opportunities in science (Luttenberger et al., 2019). In order to support students in their science development, educators, counselors, or parents need to know more about their experiences, attitudes, and the family's and social context's view on science. Here, standardized measurement instruments can be helpful.

1.2.1. Overview of measurement instruments

Ideally, a measurement instrument should meet several quality criteria. It should capture students' individual assessments of their abilities, aptitudes, values, and attitudes of their close social network (e.g., parents), as well as actual experiences. High assessment quality concerning validity and reliability as well as economic use concerning time and administration are also important criteria for a measurement instrument. Another important aspect concerns the age range for which it can be applied. As described above, attitudes and interest in science develop from an early age and may still change in adolescence (Luttenberger et al., 2018). Thus, it would be important to capture students' science development at different age ranges, from elementary education onward to middle and high school. The few existing surveys on science capital, science attitudes, and science experiences we found in our examination of the literature point to the need for the development of a new instrument. Concerns pertain to restrictions in the age range, a narrow focus in terms of content, as well as assessment quality issues and economy of use.

Most assessments on science attitudes, science capital, beliefs, and experiences focus on a narrow age range, typically young adolescents. For example, the *Test of Science-Related Attitudes* (TOSRA; stelar, 2021; Fraser, 1978, 1981; further developments Fraser & Lee, 2015) assesses attitudes towards science of 15- to 16-year-old students (grade 9–10). It is useful for capturing adolescents' interest in science activities and careers and attitudes towards science and scientists, and scientific inquiry as a way of thought. However, the TOSRA neither considers science experiences nor does it regard the value a person attaches to science; furthermore it is restricted to the age range of later adolescence. The *Relevance of Science Education* (ROSE) survey (Schreiner & Sjøberg, 2004; Sjøberg & Schreiner, 2010; Jenkins & Nelson, 2005) shows similar restrictions in age range with a focus on 15-year-old learners. Moreover, its developers noted that they did not explicitly examine constructs from a psychometric perspective (Sjøberg & Schreiner, 2010) and measurement information has not been published. With a length of 247 items, the survey also lacks economy and practicality (Aydeniz & Kotowski, 2014). With a focus on 10- to 14-year-olds, the ASPIRES survey by Archer and DeWitt (2016) captures a wider age range (but does not apply to students in elementary school). It measures different constructs like aspirations related to science, science interests outside of school, the support and aspirations of parents, attitudes to science held by peers, science experiences inside school, images of science, science self-concept, and factors related to science careers (Archer & DeWitt, 2016). However, the survey was developed for youth in school in the United Kingdom and the items have been critiqued as being less relevant or understood by youth outside of that country (Jones et al., 2021). The Student Attitudes Toward STEM Survey (S-STEM) is the only measurement instrument which addresses upper elementary as well as middle/high school students' attitudes toward STEM (Unfried et al., 2015). However, the S-STEM has shortcomings if researchers seek to examine science in general and not specific subjects, and it also fails to take into account the influence of parents on career aspirations.

As this description shows, there are various shortcomings of existing measurements. As a response to the need for valid instruments which address students' science experiences and attitudes, Jones and colleagues developed the *NextGen Scientist Survey* for middle school students which is described below (Jones et al., 2021).

1.2.2. The Next Generation Scientist Survey

The Next Generation Scientist Survey examines different factors that may influence youths' science attitudes and career aspirations including science self-efficacy/self-concept, expectancy values concerning science and a science career, and family science habitus and capital. Structural validity of the survey was examined in a recent study with more than 1000 middle school students (for details of the validation, see Jones et al., 2021). The Survey comprises four factors.

The factor *Science Achievement Value* describes how adolescents assess themselves in science as well as how others seem to perceive them. This factor includes beliefs about both science academic self-concept as well as science self-efficacy. These two constructs are overlapping and both are related to the individual's sense of competencies in science (Eccles & Wigfield, 2020). Self-efficacy is applied here as an individual's belief that they can be successful in accomplishing a task in the future, whereas academic self-concept for science refers to the beliefs about accomplishing a task at the current time (Bong & Clark, 1999).

The factor *Future Science Task Value* describes students' beliefs about the value of science in the future both as a career as well as for leisure activities. If a student does not believe that science is useful or interesting, they are not likely to see value in engaging in future science activities. To motivate students to consider a science career, students need to believe that science has future value and utility (Ertl, Luttenberger, & Paechter, 2017; Jones et al., 2021; Luttenberger, Steinlechner, Ertl, & Paechter, 2019).

The factor *Perceptions of Family Science Achievement Value* draws closely on the concept of social capital (Archer et al., 2012; Jones et al., 2021). Science career aspirations are also shaped by the perceptions a student has about whether or not the family sees science as important and relevant. This perception can contribute to the individual's belief that science is something people like them do now as well as in the future (Ertl et al., 2017). Items on this scale relate to a young person's science capital and family science habitus (Archer et al., 2015). If the family finds science interesting, the child is likely to have access to science capital, including access to others who support science learning and access to science resources. This type of perceived family values for science may provide the support that was identified by Lent and Brown (2019) as influences on science self-efficacy and outcome expectancies for future career decisions.

The factor *Science Experiences* describes former activities and experiences concerning sciences. Expectancy-value theory is built on the premise that students who have science experiences in- and out-of-school are more likely to develop interests and career aspirations for science. This experience could take the shape of access to high quality science classes, family science hobbies, or visits to science centers (Jones et al., 2021). Items in this factor also describe use of science tools (e.g., a microscope). One of the components of science capital is access to physical materials that can be used to learn science such as binoculars, a compass, or a meter stick. These materials and tools allow students to explore interests in science as well as develop skills with scientific practices. Earlier research on science hobbyists found that having access to that first telescope, bird feeder, or model building set was instrumental in the development of science interests that persisted over the life-span (Corin et al. 2017). Another form of science experiences concerns indirect experiences by reading or watching science contents with various media, by going online, or by talking about science (Jones et al., 2021).

In comparison to other measurement instruments, the NextGen Scientist Survey has various advantages. The survey addresses issues of quality, narrow content, and efficiency that have been critiques of other similar assessments. It captures important influences in socialization like parent's science values. Furthermore, it is built on expectancy-value theory that allows for researchers to extend and test the theoretical factors of expectancy values as they investigate future science task values. Structural validity was investigated by exploratory as well as confirmatory factor analyses and showed good quality of the economic-to-use questionnaire (Jones et al., 2021). In light of these advantages, it would be desirable to expand the age range to younger students.

1.3. Research questions and research design

The purpose of this research was to develop and validate the *NextGen Scientist Survey* as an instrument that can be used with elementary and middle school students to explore factors related to science interests, experiences, and career aspirations. In the sections that follow, we describe the processes we used to validate the assessment and the outcomes of this process.

As an extension of the middle school survey, a parallel version was designed for younger students in elementary education, from 8 to 11 years of age (the *NextGen Scientist Survey Elementary*). The two questionnaires for the different age groups are in the focus of this paper. The following research goals related to the surveys are investigated in two studies:

Study 1 - research goal 1: Adaptation of the middle school student *NextGen Scientist Survey* (Jones et al., 2021) for elementary school students and an investigation of the structural validity of the measurement instrument. The main research question was whether and to what extent this questionnaire is a psychometrically sound instrument for assessing factors that contribute to career aspirations for elementary school students.

Study 2 - research goal 2: Investigation whether the factorial structure is the same for both measurement instruments (elementary and middle school).

Research goal 2 includes two steps: First, the factor structure found in the instrument for elementary school students is examined to determine whether it holds true for a sample of middle school students. Second, measurement invariance between the two questionnaires and samples, the *NextGen Scientist Survey* for middle and elementary school, is investigated. The following research questions are the focus of this study: First, which types of measurement invariance can be assumed, i.e., whether the instruments allow a comparison of values across different age groups (elementary and middle school students). Second, which type of measurement invariance can be assumed for the samples of girls and boys (i.e., whether the instrument allows a comparison across genders).

2. Study 1: Development and investigation of the use of the NextGen Scientist Survey for elementary school students

2.1. Method

2.1.1. Materials

The *NextGen Scientist Survey Elementary* assessment was framed on the previously validated *NextGen Scientist Survey Middle School* (Jones et al., 2021). Experts (elementary and middle school science teachers and researchers) assessed which items of the *NextGen Science Survey Middle School* could be retained and which should be changed or deleted from the questionnaire. For a few items, the wording was slightly modified in order to be better understood or to fit better to children's perspectives (e.g., "After I finish high school, I will use science often." versus "After high school I will use science often"). One item ("My parents have explained to me that science is useful for my future") was omitted from the elementary survey.

The final questionnaire, the *NextGen Science Survey Elementary*, consists of 29 items that, in the version for middle school students, were allocated to one of four correlated scales (see Table 2 for items).

The *Science Achievement Value* scale includes nine items. Individuals assess on a 5-point Likert scale the degree to which they have confidence in their ability to do well in science, can learn science, and whether they believe they are viewed as someone who likes science by others (i.e., teachers and parents). The *Future Science Task Value* scale is comprised of three items. Individuals assess whether they will need to use science in a future job or whether they generally will use science in the future on a 5-point Likert scale. *The Perceptions of Family Science Achievement Values* scale is comprised of three items. Individuals assess the degree to which they perceive their family as finding science interesting, useful, and important on a 5-point Likert scale. The *Science Experiences* scale is made up of 14 items which ask on a 5-point scale how often a person has carried out certain activities related to science. Of these, 10 items refer to science activities that include hands-on activities and direct experiences (tangible science experiences), while 4 items measure a more theoretically-oriented engagement with science (intangible science experiences) such as reading books, talking about science with other people, or going online to learn science.

2.1.2. Sample

Overall 327 students in grades 3–5 (age range 8–11) filled in the questionnaire (grade 3: $n = 18$, 5.5 %; 4: $n = 163$, 49.85 %; 5: $n = 144$, 44.04 %; two missing values, 0.61 %).

As the survey is a self-respondent measure and a minimum of reading skills was required, the survey targeted children from grade 3 onward and 8-years and older. As the survey for middle school students targeted children in grade 6, students up to grade 5 were chosen for the elementary school survey.

The children in the study identified as male ($n = 141$, 43.12 %), female ($n = 178$, 54.43 %) (8 missing values, 2.45 %); Caucasian ($n = 69$, 21.10 %), African American ($n = 130$; 39.76 %), Hispanic ($n = 75$, 22.94%), American Indian/Alaskan ($n = 5$; 1.53 %), other ($n = 32$, 9.79) and 16 missing values (4.89%). Recruitment of students was from after school programs that served all students, but the programs reached out to diverse communities to make sure all students have access to after school programs.

The study was performed in accordance with the American Psychological Association's Ethics Code and the Declaration of Helsinki. Participants were contacted in person at the afterschool program and asked to complete the survey at the program. For each participant, consent to participate was given by their parent and assent was garnered from the child. The study was approved by the first author's Institutional Review Board for Research With Human Subjects.

2.1.3. Data analyses

For investigation of the structural validity of the *NextGen Scientist Survey Elementary* two analyses were conducted. Probability analyses for item difficulty and confirmatory factor analyses for investigation of the factor structure were carried out. Four theoretical models assuming different factor structures were investigated by confirmatory factor analyses (CFA). For the purpose of scaling the factors, in each model and for each factor, one factor loading was fixed to one. Several model fit indices were used to identify the most appropriate measurement model, namely Akaike's Information Criterion (AIC), sample size adjusted Bayesian Information Criterion (BIC), χ^2 values, comparative fit index (CFI), root mean square error of approximation (RMSEA), and standardized root mean squared residual (SRMR). Thus it could be investigated which of the competing models fits the data best and is the most appropriate model to retain for future use. The inspection of fit indices followed recommendations by Strijbos and colleagues (2021, p. 6; similar recommendations also in Paechter et al., 2013; Papousek et al., 2012): "Standardized Root Mean-square Residual (SRMR) and Root Mean Square Error of Approximation (RMSEA) below .10 is considered adequate fit and below.05 an excellent fit, and Comparative Fit Index (CFI) scores above.90 indicate adequate fit and above .95 excellent fit." For the sake of completeness, χ^2-indices were recor ded. However, they should be regarded cautiously since the χ^2 statistic becomes increasingly unreliable in large sample sizes > 250 (Putnick & Bornstein, 2016; Strijbos et al., 2021). All analyses were conducted using SPSS 26 and R 3.4.4 plus lavaan 0.6–6 (Rosseel, 2012).

2.2. Results

Item difficulty analysis showed that only one item was outside the range.20 $\leq p_i \leq$ 0.80. Since this questionnaire is one of two partly parallel forms and since the deviation from the desired values was quite small ($p = .83$) the item remained in the data set.

Based on the theoretical structure of the questionnaire and on the previous empirical analysis of the *NextGen Scientist Survey Middle School* for students in grades six to eight (Jones et al., 2021), four theoretical models were investigated (see Table 1): Model 1 is a baseline model in which all items were constrained to load onto one single general factor. The purpose of this model was to confirm that there was some level of distinctiveness among the items in that there were multiple underlying latent factors. Model 2 has all items loading on one of two factors based on their answer format (assessing the frequency of activities and assessing the degree of acceptance to an item, respectively). The 15 items with a 5-point Likert scale indicating agreement are assumed to load on one factor, they comprise items of the *Science Achievement Value* scale, the *Future Science Task Value* scale, the *Perceptions of Family Science Achievement Values* scale. The other assumed factor comprises the 14 items of the *Science Experiences* scale with a 5-point scale indicating frequencies. Model 3 is a 4-factor model that represents the original structure of the *NextGen Scientist Survey* for middle school students with four scales. Model 4 is a 5-factor model. It represents the original structure of the *NextGen Scientist Survey* as previously validated (Jones et al., 2021) with the *Science Achievement Value* scale, the *Future Science Task Value* scale, and the *Perceptions of Family Science Achievement Values* scale. Then the *Science Experiences* scale was split up into two factors to allow for more differentiation: tangible science experiences versus intangible science experiences.

The model comparison in Table 1 shows unsatisfactory fit indices for model 1, 2, and 3. Model 4 fits the data best. RMSEA is within the desired boundary as well as SRMR. CFI has an adequate fit value.

The left side of Table 2 shows the final solution for the 5-factor model for the sample of elementary students (model 4). The items of the questionnaire show good to acceptable selectivity ($r_{it} \geq 0.3$; Field et al., 2012) with values of .377 $\leq r_{it} \leq$ 0.720 (see Table 2). With one exception, the five factors also show satisfactory to good reliability ($\alpha \geq 0.7$; Field et al., 2012) with Cronbach's α values of .725 $\leq \alpha_t \leq$ 0.921. Although the factor Perceptions of Family Science Achievement Value shows a reliability slightly below the cutoff ($\alpha = 0.695$), we decided to keep it as part of the questionnaire. First, its reliability is very close to the arbitrarily set cutoff. And second, Cronbach's α tends to underestimate the reliability of scales with a smaller number of items (Field et al., 2012) and the factor of concern does consist of three items. All item to factor loadings of the NextGen Scientist Survey Elementary are significant ($p < .05$) and substantial ($\beta > 0.4$; Field et al., 2012).

Table 3 shows descriptive statistics for the five scales in the sample of the elementary school children and the correlations between the scales.

Table 1
Comparison of CFA-models of the *NextGen Scientist Survey Elementary*.

Model	*n*	*AIC*	*BIC*	χ^2	*df*	χ^2/df	*CFI*	*RMSEA*	*SRMR*
1	327	26834.960	27164.687	1352.651	377	3.588	.683	.089	.084
2	327	26465.339	26798.856	981.030	376	2.609	.803	.070	.066
3	327	26201.209	26553.675	706.899	371	1.905	.891	.053	.058
4	327	26148.700	26516.326	646.390	367	1.761	.909	.048	.053

3. Study 2: Replication of the factorial structure of the *NextGen Scientist Survey Elementary* for the survey for middle school students and test for invariance

3.1. Method

3.1.1. Materials

The original *NextGen Scientist Survey Middle School* (Jones et al., 2021) was comprised of 30 items of which 29 are parallel to the *NextGen Scientist Survey Elementary*. These 29 items were used for the following analyses.

3.1.2. Sample

The original survey was given nationwide to youths in grades 6–8 by teachers who volunteered to participate in the study in response to email and listserv requests. Teachers gave the assessment during science classes. Overall, 889 students, in grades 6–8 (age range 11–14 years of age), completed the questionnaire (grade 6: $n = 32$, 3.60 %; 7: $n = 658$, 74.02 %; 8: $n = 199$, 22.38 %). The children identified as male ($n = 450$, 50.62%), female ($n = 439$, 49.38 %); Caucasian ($n = 538$, 60.52 %), African American ($n = 62$, 6.97%), Hispanic ($n = 107$, 12.04 %), American Indian/Alaskan ($n = 27$, 3.04 %), Asian ($n = 39$, 4.39 %), and Other ($n = 116$, 13.05 %).

The study was performed in accordance with the American Psychological Association's Ethics Code and the Declaration of Helsinki. For each participant, consent to participate was given by their parent and assent was garnered from the adolescent. The study was approved by the first author's Institutional Review Board for Research With Human Subjects.

3.1.3. Data analyses

First, the fit of a 5-factor structure of *NextGen Science Survey Middle School* data to the *NextGen Scientist Survey Elementary* was investigated. Item difficulty analyses and CFA were carried out.

Second, measurement invariance was investigated. In various instances, it would be desirable to use the *NextGen Scientist Survey* for comparisons, such as to observe children's development of attitudes towards sciences over a longer time range or to compare different age groups. Measurement invariance means that the same constructs are measured across the different age groups or across time and that comparisons reflect true differences between the groups (Ma & Qin, 2021; Putnick & Bornstein, 2016). The methodological approach for the determination of measurement invariance will be explained below; it followed research studies with similar aims (e.g., Strijbos et al., 2021; Schmitt et al., 2011).

CFA allows for tests of different forms of invariance which can be ordered with regard to strictness in the sense of dimensions in which invariance can be assumed. In the present study, data were tested for three different forms of invariance, configural, metric, and scalar invariance.

Configural invariance means that the number of factors and the configuration of items belonging to a factor is invariant across groups (Gregorich, 2006). However, with configural invariance factor-intercorrelations and factor weights may differ across groups; therefore, it does not allow quantitative comparisons between groups (Ma & Qin, 2021; Strijbos et al., 2021). Usually, the configural invariance model serves as a baseline model for further tests on invariance (Ma & Qin, 2021). If configural invariance can be confirmed, usually metric invariance is examined. Metric invariance (weak factorial invariance) tests whether factors have the same meaning across groups, i.e., whether factor weights are equal across groups. The next step usually is an examination of scalar invariance. This is desirable, because scalar invariance (strong factorial invariance) means that factor loadings and both item and factor intercepts are equal across groups. Scalar invariance means that in all groups, there is the same relationship between the items and the measured constructs. Scalar invariance also allows comparisons between age groups (like elementary and middle school children) and individuals of different age groups or between genders and the use of statistical methods such as t-test or ANOVA to test group differences of means (Ma & Qin, 2021).

3.2. Results

3.2.1. Structural validity of the NextGen Science Survey Middle school

Analysis of item difficulties showed that all items were inside the range $.20 \leq p_i \leq 0.80$ and could be used for factor analyses. As a second step, it was investigated whether the 5-factor solution of the *NextGen Science Survey Elementary* also fits to the data of the *NextGen Science Survey* for middle school. The factors were the *Science Achievement Value* scale, the *Future Science Task Value* scale, the *Perception of Family Science Achievement Value*, and two factors concerning science experiences, the *Intangible Science Experiences* scale and the *Tangible Science Experiences* scale. CFA showed good fit indices with *RMSEA* and *SRMR* below .05 and *CFI* close to .95 (see Table 4).

Further information of the CFA (item probabilities, item selectivities, Cronbach's α, item to factor loadings) is given on the right side of Table 2. Descriptive statistics for the factors and bivariate correlations between factors are shown in Table 5.

3.2.2. Measurement invariance across age groups of elementary and middle school students

For the examination of invariance across both data sets, the samples of elementary school students with $n = 327$ and of middle school students with $n = 889$ were employed ($N = 1216$). The CFA 5-factor-model performed well in this pooled dataset of children and adolescents from grades 3–8 ($\chi^2 = 1252.76$, $df = 367$, $p < .01$, CFI = 0.941, RMSEA = 0.045, SRMR = 0.040).

Invariance testing across the elementary and middle school samples was performed (Table 6). Results showed that configural and metric invariance can be assumed. This means that the factorial structure (configural invariance) as well as the unstandardized loadings (metric invariance) are equal across groups. In other words: the latent constructs have the same meaning within the two subsamples. However, full scalar invariance could not be assumed. For scalar invariance to hold, intercepts have to be equal across groups and configural and metric invariance have to hold as well. Therefore, we tested for partial scalar invariance. This involved freeing the constraints for individual parameters to determine whether the measure as a whole cannot be compared across groups, or whether it is only certain items that are responsible for the decline in measurement fit between steps of invariance testing (Vandenberg & Lance, 2000).

Partial scalar invariance was obtained by letting three intercepts vary across groups (items 9 – factor *Science Expectancy Value*; 18 and 20 – factor *Tangible Science Experiences*). According to Dimitrov (2010)

Table 2
NextGen Scientist Elementary and *NextGen Scientist Survey Middle School*, item probability (p_i), item selectivity (r_{it}), Cronbach's α_t for each factor and item loadings β.

NextGen Scientist Sur-vey Elementary (Study 1)				*NextGen Scientist Survey Middle School (Study 2)*		
p_i	α_t/ r_{it}	β_i	Item	p_i	α_t/ r_{it}	β_i
	.859		*Science Achievement Value*		.921	
.65	.637	.706	I think I am good at science.	.56	.793	.834
.58	.667	.740	I know a lot about science.	.51	.791	.822
.62	.595	.662	I learn new science topics easily.	.56	.753	.782
.68	.551	.605	I am good at using science tools.	.62	.566	.595
.75	.671	.726	I know I can do well in science.	.63	.723	.765
.62	.629	.682	My friends think I'm good in science.	.54	.751	.783
.67	.560	.612	My teacher sees me as someone who likes science.	.54	.765	.789
.68	.590	.656	My parents see me as someone who likes science.	.49	.745	.789
.64	.418	.463	An adult has encouraged me to study science.	.50	.603	.644
	.827		*Future Science Task Value*		.865	
.54	.650	.728	When I am older I will need science for my job.	.57	.714	.768
.54	.720	.799	I would like to have a job that uses science.	.48	.774	.871
.55	.686	.828	After high school I will use science often.	.52	.745	.837
	.695		*Perceptions of Family Science Achievement Value*		.798	
.63	.543	.691	My parents think science is very interesting.	.53	.653	.764
.69	.507	.700	My family thinks it is important for me to learn science.	.62	.593	.732
.58	.485	.578	My family knows a lot about science.	.47	.683	.772
	.796		*Tangible Science Experiences (hands-on activities)*		.818	
.56	.448	.483	How many times have you gone to a museum when not in school?	.57	.430	.450
.54	.487	.516	How many times have you gone on a nature walk when not in school?	.63	.435	.467
.28	.388	.462	How many times have you read a map to find your way when not in school?	.32	.453	.498
.36	.451	.512	How many times have you planted seeds and watched them grow when not in school?	.35	.528	.601
.40	.497	.553	How many times have you used binoculars or a telescope when not in school?	.42	.590	.668
.38	.448	.509	How many times have you used a thermometer when not in school?	.43	.532	.589
.64	.525	.593	How many times have you used a ruler or measuring tape when not in school?	.64	.505	.564
.63	.377	.427	How many times have you built or taken things apart when not in school	.36	.431	.478
.44	.524	.630	How many times have you done experiments or used science kits when not in school	.33	.534	.615
.67	.531	.561	How many times have you collected rocks or shells when not in school?	.50	.547	.618
	.725		*Intangible Science Experiences*		.795	
.51	.472	.565	How many times have you watched science programs on TV when not in school?	.47	.569	.646
.42	.519	.617	How many times have you read a book or magazine about science when not in school?	.31	.634	.708
.41	.534	.643	How many times have you gone online to look up information about science when not in school?	.28	.656	.748
.52	.531	.696	How many times have you talked about science with others when not in school?	.41	.567	.714

Note. Left side of Table 2 shows values for the elementary school sample used in Study 1, right side the values for same analyses in Study 2.

Table 3
NextGen Scientist Survey Elementary, mean values (M), standard deviations (SD), and bivariate correlations between factors.

Scale	*M*	*SD*	(2)	(3)	(4)	(5)
Science Achievement Value (1)	3.62	0.761	.512	.562	.387	.509
Future Science Task Value (2)	3.19	1.063		.454	.310	.432
Perceptions of Family Science Achievement Value (3)	3.53	0.847			.359	.440
Tangible Science Experiences (4)	1.47	0.661				.599
Intangible Science Experiences (5)	1.40	0.837				

partial invariance is acceptable for practical use as long as the number of freed parameters does not exceed 20 %. For the present study 3 out of 53 constraints were freed (5.66 %), thus staying well within this limit. This means that, since partial scalar invariance does hold, the factor means can be compared across groups.

3.2.3. *Measurement invariance for gender*

It may also be desirable to use the *NextGen Science Survey* for comparisons between females and males. Therefore, the measurement invariance for gender was investigated. Again configural, metric, and scalar invariance were examined. As Table 6 and the comparison of models by Δ CFI show, full scalar invariance can be assumed for gender. This allows a comparison between females and males and the use of statistical methods such as t-test or ANOVA to test for group differences of means are allowed (Li et al., 2016; Ma & Qin, 2021).

4. Discussion

The studies presented in this paper describe the first validation studies for the *NextGen Scientist Survey Elementary* with assessments of structural validity as well as measurement invariance across elementary and middle school samples as well as across genders.

4.1. *Structural validity of the NextGen Scientist Surveys*

Using CFA, the *NextGen Scientist Survey Elementary* was shown to be psychometrically sound for five correlated factors shown to influence children's career aspirations: *Science Achievement Value*, *Future Science Task Value*, *Family Science Achievement Value*, *Tangible Science Experiences*, *Intangible Science Experiences*. CFA found five factors, three of them measure self-assessments and assessments concerning science, two measure experiences in science. This structure was also confirmed for the *NextGen Scientist Survey Middle School*.

The *Science Achievement Value* scale shows high reliability values in both samples. It includes items that express self-assessments (e.g., I think I am good at science) as well as items in which the child/adolescent describes how others see them (e.g., My teacher/parents sees me as someone who likes science). For both age groups, elementary and middle school children, self-assessment and the assumed assessment of others were highly correlated. The factor *Future Science Task Value*, also with high reliability, assesses perceptions of the need to use science in a future job and to use science in general in the future. The *Perceptions of Family Science Achievement Value* scale measures beliefs about how interesting and valuable science is to parents and family. Reliability values were high in both samples. If the family finds science interesting, now and in the future, the child is likely to have access to science capital, including access to others who support science learning and access to science resources.

The factor solutions found for the NextGen Scientist Survey for elementary and middle school students differ slightly but favorably from the first factor solution found for middle school students. The differences

Table 4
Fit indices for CFA-model for the *NextGen Science Survey Middle School*.

n	*AIC*	*BIC*	χ^2	*df*	χ^2/df	*CFI*	*RMSEA*	*SRMR*
889	67034.86	67499.50	1058.600	367	2.884	.942	.046	.043

Table 5
NextGen Scientist Survey Middle School, mean values (M), standard deviations (SD), and bivariate correlations between factors.

Scale	*M*	*SD*	(2)	(3)	(4)	(5)
Science Achievement Value (1)	3.20	0.881	0.624	.603	.390	.557
Future Science Task Value (2)	3.09	0.995		.520	.254	.436
Perceptions of Family Science Achievement Value (3)	3.16	0.846			.284	.410
Tangible Science Experiences (4)	1.37	0.685				.588
Intangible Science Experiences (5)	1.10	0.877				

that were found relate to science experiences. The items on *Science Experiences* split into two factors for the elementary students sample: hands-on, *Tangible Experiences* and *Intangible Experiences*. This factor solution also fits well in the sample of middle school students even though in the original survey and its validation, all experience items formed only one factor (Jones et al., 2021). Reliability values for the science experience factors were high in both samples. Also, the distinction between the two kinds of experiences allows a more precise description as in reality tangible science experiences like visiting museums, using tools, etc. do not need to be related to more intangible experiences like acquiring science knowledge by talks or by using media.

4.2. Measurement invariance: Implication for use of the NextGen Scientist Surveys

The present research could confirm partial scalar invariance for the elementary and middle school students samples thus allowing comparison of factor means across groups (following requirements described by Dimitrov, 2010). In such cases of invariance also effect sizes like Cohen's d (1988) can be reported (Putnick & Bornstein, 2016) to assess the importance of group differences. Furthermore, full scalar invariance across genders could be confirmed thus allowing comparisons between girls and boys.

These results are important for practical use of the *NextGen Scientist Survey*. In many educational instances it might be interesting to compare samples of different ages or to investigate the influence of gender, for example when educational programs are carried out. Having surveys that measure the same factors across different age groups can allow researchers to track the different factors across time and to examine developmental changes. In addition, the survey can inform practitioners in the development of targeted programming for youth.

According to our examination of the literature on science capital questionnaires, rigid testing of validity and measurement invariance has only been carried out for the S-STEM (Unfried et al., 2015) and now for the NextGen Scientist Survey (see 1.1.1 Overview of measurement instruments). However, both instruments have a slightly different focus. The NextGen Scientist Survey, therefore, would be a suitable instrument when science in general and also parents' influence on their children's science attitudes are to be considered.

5. Conclusions and limitations

To conclude, our research efforts have tested and found acceptable reliability and validity for a new elementary level assessment of science capital, science attitudes, and science experiences, factors that contribute to career aspirations (Blanchard & Lichtenberg, 2003). Furthermore, the elementary school version of the *NextGen Scientist Survey* can be used in parallel to the middle school version to allow for assessments across educational levels and age. The ability to measure factors which contribute to career aspirations in elementary students is important because youths at this age are already forming and establishing ideas about future careers. In order to provide the support necessary to sustain interest in science from elementary ages to middle school (and beyond) there is a need to have a longitudinal understanding of the resources to which youths have access. There is much concern about the "leaky STEM pipeline" (Makarova, Aeschlimann, & Herzog, 2016) and the loss of interest in STEM (science, technology, engineering, mathematics) disciplines as they get older. The development of this survey provides researchers with the opportunity to examine factors that contribute to career aspirations that can be measured through the critical transitional years when youths begin to lose interest in science. Therefore, the parallel forms of this instrument, for elementary and middle school youths, can allow for documentation of how youths' attitudes toward science can change over time through the lens of family science capital and family science habitus as well as how female and male students vary in their science attitudes and experiences.

Altogether, the two versions of the NextGen Scientist Survey can be described as questionnaires with well-grounded structural validity and they meet important quality characteristics concerning measurement invariance across different age groups and genders. Thus the surveys can be effectively used. However, future research to further explore reliability by assessing the same children twice within a short period of time would be desirable. Additional validity supportive evidence might come from studies investigating cross-age differences in larger samples and from more distant age categories. It would be also worthwhile to assess reliability and validity within the data set as well as beyond the data set and to set up new studies with measures of construct validity, discriminant and convergent validity. This research will help refine the survey and, perhaps more importantly, advance our understanding of factors that contribute to future science task value.

Table 6
Fit indices for tests on invariance for age groups and gender.

Age groups	$\Delta\ \chi^2$	*Δ df*	*CFI*	*Δ CFI*	*RMSEA*	*Δ RMSEA*	*SRMR*	*Δ SRMR*
baseline configural			.936		.047		.046	
1 metric	41.524 *	24	.934	-0.002	.046	-0.001	.048	.002
2 scalar	355.566 * *	24	.912	-0.022	.053	.007	.052	.004
3 partial scalar	138.147 * *	21	.927	-0.007	.048	.002	.050	.002
Gender	$\Delta\ \chi^2$	*Δ df*	*CFI*	*Δ CFI*	*RMSEA*	*Δ RMSEA*	*SRMR*	*Δ SRMR*
baseline configural			.937		.046		.045	
1 metric	28.455	24	.937	.000	.046	.000	.047	.002
2 scalar	123.166 * *	24	.930	-0.007	.047	.001	.048	.001

Note. * * = $p < .01$, * = $p < .05$.

It is possible that there may have been some selection bias present in the samples that volunteered to participate in the study. The percentage of ethnicities in the elementary school sample differs from the percentage in the state in which the research was carried out. Also, a larger size for the elementary student sample would have been desirable. Despite these limitations, this study gives a first insight into the promising development of an instrument to capture the science capital, attitudes, and future science task value of children. Further sampling is needed to determine important characteristics of the questionnaire, for example invariance between genders. Additional research is needed to establish construct, convergent, and discriminant validity, criterion-related validity, and test-retest reliability.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. (1614468).

References

Archer, L., Dawson, E., DeWitt, J., Seakins, A., & Wong, B. (2015). "Science capital": A conceptual, methodological, and empirical argument for extending Bourdieusian notions of capital beyond the arts. *Journal of Research in Science Teaching, 52*(7), 922–948. https://doi.org/10.1002/tea.21227

Archer, L., & DeWitt, J. (2016). *Understanding young people's science aspirations: How students form ideas about 'becoming a scientist.'*. London: Routledge.

Archer, L., DeWitt, J., Osborne, J., Dillon, J., Willis, B., & Wong, B. (2012). Science aspirations, capital, and family habitus: How families shape children's engagement and identification with science. *American Educational Research Journal, 49*(5), 881–908. https://doi.org/10.3102/0002831211433290

Aydeniz, M., & Kotowski, M. R. (2014). Conceptual and methodological issues in the measurement of attitudes towards science. *Electronic Journal of Science Education, 18* (3), Article 8531 (Article) ⟨https://ejrsme.icrsme.com/article/view/8531⟩.

Blanchard, C. A., & Lichtenberg, J. W. (2003). Compromise in career decision making: A test of Gottfredson's theory. *Journal of Vocational Behavior, 62*, 250–271. https://doi.org/10.1016/S0001-8791(02)00026-X

Bong, M., & Clark, R. E. (1999). Comparison between self-concept and self-efficacy in academic motivation research. *Educational Psychologist, 34*(3), 139–153. https://doi.org/10.1207/s15326985ep3403_1

Claussen, S., & Osborne, J. (2013). Bourdieu's notion of cultural capital and its implications for the science curriculum. *Science Education, 97*(1), 58–79. https://doi.org/10.1002/sce.21040

Corin, E., Jones, M. G., Andre, T., & Childers, G. (2017). Characteristics of lifelong science learners: An investigation of STEM hobbyists. *International Journal of Science Education, Part B, 8*(1), 53–75. https://doi.org/10.1080/21548455.2015.1118664

DeWitt, J., Archer, L., & Moote, J. (2019). 15/16-year-old students' reasons for choosing and not choosing physics at a level. *International Journal of Science and Mathematics Education, 17*, 1071–1087. https://doi.org/10.1007/s10763-018-9900-4

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development, 43*, 121–149. https://doi.org/10.1177/0748175610373459

Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation (Article) *Contemporary Educational Psychology, 61*(4), Article 101859. https://doi.org/10.1016/j.cedpsych.2020.101859.

Ertl, B., Luttenberger, S., & Paechter, M. (2017). The impact of gender stereotypes on the self-concept of female students in STEM subjects with an under-representation of females (Article) *Frontiers in Psychology, 8*, Article 703. https://doi.org/10.3389/fpsyg.2017.00703.

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. London: Sage.

Fraser, B. J. (1978). Development of a test of science-related attitudes. *Science Education, 62*(4), 509–515. https://doi.org/10.1002/sce.3730620411

Fraser, B. J. (1981). *TOSRA: Test of science-related attitudes: Handbook*. Victoria: Australian Council for Educational Research.

Fraser, B. J., & Lee, S. U. (2015). Use of test of science related attitudes (TOSRA) in Korea. In C. S. Khine (Ed.), *Attitude measurements in science education: Classic and contemporary approaches* (pp. 293–308). Charlotte, NC: IAP.

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory analysis framework, 11 *Medical Care, 44*(3), 78–94. https://doi.org/10.1097/01.mlr.0000245454.12228.8f.

Jacobs, J. E., & Bleeker, M. M. (2004). Girls' and boys' developing interests in math and science: Do parents matter. *New Directions for Child and Adolescent Development, 106*, 5–21. https://doi.org/10.1002/cd.113

Jenkins, E. W., & Nelson, N. W. (2005). Important but not for me: Students‘ attitudes towards secondary school science in England. *Research in Science & Technological Education, 23*, 41–57. https://doi.org/10.1080/02635140500068435

Jones, M. G., Chesnutt, K., Ennes, M., Mulvey, K. L., & Cayton, E. (2021). Understanding science career aspirations: Factors predicting future science task value. *Journal of Research in Science Teaching.* https://doi.org/10.1002/tea.21687

Jones, M. G., Childers, G., Corin, E., Chesnutt, K., & Andre, T. (2019). Free choice learning and STEM career choice. *International Journal of Science Education, Part B, 9* (1), 29–39. https://doi.org/10.1080/21548455.2018.1534024

Jones, M. G., Ennes, M., Weedfall, D., Chesnutt, K., & Cayton, E. (2021). The development and validation of a measure of science capital, habitus, and future science interests. *Research in Science Education*. https://doi.org/10.1007/s11165-020-09916-y

Kessels, U., Rau, M., & Hannover, B. (2006). What goes well with physics? Measuring and altering the image of science. *British Journal of Educational Psychology, 76*, 761–780. https://doi.org/10.1348/000709905×59961

Lent, R. W., & Brown, S. D. (2019). Social cognitive career theory at 25: Empirical status of the interest, choice, and performance models (Article) *Journal of Vocational Behavior, 115*, Article 103316. https://doi.org/10.1016/j.jvb.2019.06.004.

Luttenberger, S., Steinlechner, P., Ertl, B., & Paechter, M. (2019). It takes more than one swallow to make a summer: Measures to foster girls' and women's pathways into STEM (Article) *Frontiers in Psychology, 10*, Article 1844. https://doi.org/10.3389/fpsyg.2019.01844.

Luttenberger, S., Wimmer, S., & Paechter, M. (2018). Spotlight on math anxiety. *Psychology Research and Behavior Management, 11*, 311–322. https://doi.org/10.2147/PRBM.S141421

Ma, Y., & Qin, X. (2021). Measurement invariance of information, communication and technology (ICT) engagement and its relationship with student academic literacy: Evidence from PISA 2018. *Studies in Educational Evaluation, 68*, Article 100982. https://doi.org/10.1016/j.stueduc.2021.100982

Makarova, E., Aeschlimann, B., & Herzog, W. (2016). Why is the pipeline leaking? Experiences of young women in STEM vocational education and training and their adjustment strategies. *Empirical Research in Vocational Education and Training, 8*(2), 1–18. https://doi.org/10.1186/s40461-016-0027-y

Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Arens, A. K. (2019). The murky distinction between self-concept and self-efficacy: Beware of lurking jingle-jangle fallacies. *Journal of Educational Psychology, 111*, 331–353. https://doi.org/10.1037/edu0000281

Newall, C., Gosalkorale, K., Walker, E., Forbes, A., Highfield, K., & Sweller, N. (2018). Science education: Adult biases because of the child's gender and gender stereotypicality. *Contemporary Educational Psychology, 55*, 30–41. https://doi.org/10.1016/j.cedpsych.2018.08.003

Nugent, G., Barker, B., Welch, G., Grandgenett, N., Wu, C., & Nelson, C. (2015). A model of factors contributing to STEM learning and career orientation. *International Journal of Science Education, 37*(7), 1067–1088. https://doi.org/10.1080/09500693.2015.1017863

Wimmer, S., Paechter, M., Lackner, H., & Papousek, I. (2020). Effects of self-concept on restoring a positive motivational state after an exam-like situation. *Mind Brain and Education, 14*, 5–15. https://doi.org/10.1111/mbe.12222

OECD (2020). PISA 2024 Strategic vision and direction for science. ⟨https://www.oecd.org/pisa/publications/PISA-2024-Science-Strategic-Vision-Proposal.pdf⟩ Accessed February 10, 2021.

Paechter, M., Rebmann, K., Schloemer, T., Mokwinski, B., Hanekamp, Y., & Arendasy, M. (2013). Development of the oldenburg epistemic beliefs questionnaire (OLEQ), a German questionnaire based on the epistemic belief inventory (EBI). *Current Issues in Education, 16*(1), 100–110.

Papousek, I., Ruggeri, K., Macher, D., Paechter, M., Heene, M., Weiss, E. M., … Freudenthaler, H. H. (2012). Psychometric evaluation and experimental validation of the statistics anxiety rating scale. *Journal of Personality Assessment, 94*(1), 82–91. https://doi.org/10.1080/00223891.2011.627959

Oon, P.-T., & Subramaniam, R. (2013). Factors influencing Singapore students' choice of physics as a tertiary field of study: A Rasch analysis. *International Journal of Science Education, 35*(1), 86–118. https://doi.org/10.1080/09500693.2012.718098

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Rodriguez-Planas, N., & Nollenberger, N. (2018). Let the girls learn! It is not only about math … it's about gender social norms. *Economics of Education Review, 62*, 230–252. https://doi.org/10.1016/j.econedurev.2017.11.006

Rosseel, Y. (2012). lavaan: An R Package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Schmitt, N., Golubovich, J., & Leong, F. T. L. (2011). Impact of measurement invariance on construct correlations, mean differences, and relations with external correlates: An illustrative example using Big Five and RIASEC measures. *Assessment, 18*(4), 412–427. https://doi.org/10.1177/1073191110373223

Schreiner, C., & Sjøberg, S. (2004). Sowing the seeds of ROSE: Background, rationale, questionnaire development and data collection for ROSE (The Relevance of Science Education): A comparative study of students' views of science and science education. *Acta Didactica, 4*, 1–120.

Sjøberg, S., & Schreiner, C. (2010). The ROSE project: An overview and key findings. Oslo: University of Oslo. ⟨https://roseproject.no/network/countries/norway/eng/nor-Sjoberg-Schreiner-overview-2010.pdf⟩ Accessed May 27, 2021.

stelar. Australian Council for Educational Research (2021). Instruments. Test of Science Related Attitudes (TOSRA). ⟨http://stelar.edc.org/instruments/test-science-related-attitudes-tosra⟩ Accessed May 27, 2021.

Strijbos, W., Pat-El, R., & Narciss, S. (2021). Structural validity and invariance of the feedback perceptions (Article) *Studies in Educational Evaluation, 68*, Article 100980. https://doi.org/10.1016/j.stueduc.2021.100980.

UNESCO. (2017). *Cracking the code: Girls' and women's education in science, technology, engineering and mathematics (STEM)*. Paris: UNESCO.

Unfried, A., Faber, M., Stanhope, D. S., & Wiebe, E. (2015). The development and validation of a measure of student attitudes toward science, technology, engineering, and math (S-STEM). *Journal of Psychoeducational Assessment, 33*(7), 622–639. https://doi.org/10.1037/a001860

Van Tuijl, C., & van der Molen, J. H. W. (2015). Study choice and career development in STEM fields: An overview and integration of the research. *International Journal of Technology and Design Education, 26*(2), 159–183. https://doi.org/10.1007/s10798-015-9308-1

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4–70. https://doi.org/10.1177/109442810031002

Watt, H. M. (2004). Development of adolescents' self-perceptions, values, and task perceptions according to gender and domain in 7th through 11th grade Australian students. *Child Development, 75*(5), 1556–1574. https://doi.org/10.1111/j.1467-8624.2004.00757.x

Watt, H. M. G., Bucich, M., & Dacosta, L. (2019). Adolescents' motivational profiles in mathematics and science: Associations with achievement striving, career aspirations and psychological wellbeing (Article) *Frontiers in Psychology, 10*, Article 990. https://doi.org/10.3389/fpsyg.2019.00990.

Wentzel, K. R., & Wigfield, A. (2007). Motivational interventions that work: Themes and remaining issues. *Educational Psychologist, 42*(4), 261–271. https://doi.org/10.1080/00461520701621103

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology, 25*(1), 68–81.