



# Scenario Evaluation with Relevance and Interest (SERI): Development and Validation of a Scenario Measurement Tool for Context-Based Learning

Jingoo Kang<sup>1</sup>  · Tuula Keinonen<sup>1</sup> · Shirley Simon<sup>2</sup> · Miia Rannikmäe<sup>3</sup> ·  
Regina Soobard<sup>3</sup> · Inês Direito<sup>4</sup>

Received: 9 May 2018 / Accepted: 30 September 2018 / Published online: 22 October 2018  
© The Author(s) 2018

## Abstract

Context-based learning (CBL) approaches have been recommended and expanded in science education to make science more relevant to students by connecting science content with students' daily life. Subsequently, in order to implement CBL at school, a group of scenarios has been produced by several stakeholders. However, there is a lack of resources to measure effectively what makes a good scenario. Thus, this study aims to develop and validate a scenario evaluation instrument to examine students' perspectives on science career-related scenarios through the lens of relevance and interest. For this purpose, 25 science career-related scenarios and a measurement tool, Scenario Evaluation with Relevance and Interest (SERI), were developed by a team of researchers for the EU funded MultiCO project. Then, lower secondary school students from three different countries, Estonia, Finland, and the UK, were asked to respond to the newly developed instrument after reading the scenarios, and their responses were analyzed by factor analyses and multivariate analysis of variance. According to the result, this instrument has good construct validity and reliability. However, it indicates one issue of discriminant validity between two factors, individual dimension and societal dimension. Also, significant gender differences were found in the Estonian sample regarding students' perspectives on the scenarios. Possible interpretations of the results and implications of the suggested measurement tool are discussed.

---

✉ Jingoo Kang  
jingoo.kang@uef.fi

<sup>1</sup> School of Applied Educational Science and Teacher Education, University of Eastern Finland, Yliopistokatu 2, FI-80100 Joensuu, Finland

<sup>2</sup> Institute of Education, University College London, Gower Street, London WC1E 6BT, UK

<sup>3</sup> Centre for Science Education, University of Tartu, Vanemuise 46-226, Tartu 51014, Estonia

<sup>4</sup> Centre for Engineering Education, University College London, Gower Street, London WC1E 6BT, UK

**Keywords** Context-based learning · Interest · Measurement tool · Relevance · Scenario evaluation

## Introduction

In the science education community, students' negative trends in science interest have been an issue in recent decades especially at the secondary school level. As interest has been revealed as a sound predictor of students' career aspirations in their future (Kang & Keinonen, 2017; Schoon, 2001; Tai, Liu, Maltese & Fan, 2006), young pupils' decreasing interest in science has affected students' low enrolment in advanced science courses in upper secondary school and college (Fouad et al., 2010; Simon & Osborne, 2010). It has been indicated that one of the major reasons giving rise to this phenomenon has been that science at school is irrelevant to students' everyday life (Gilbert, 2006; Holbrook, 2008; Stuckey, Hofstein, Mamlok-Naaman & Eilks, 2013). This issue has resonated so much with educators, policymakers, and researchers that many projects have been initiated and conducted in order to make school science more relevant to students' daily life, with the aim of drawing students' attention to science and STEM (Science, Technology, Engineering, and Mathematics) careers. One such attempt is the context-based science curricula movement, which was initiated in the 1970s and developed with various forms (Taconis & den Brok, 2016). In particular, since contexts have been designed to relate the science content to students' life, such as socio-scientific issues, context-based learning (CBL) in science education is unique as it has tried to develop and use scenarios relating to the contexts (Lubben, Bennett, Hogarth & Robinson, 2004; Pilot & Bulte, 2006). Needless to say, a variety of scenarios has been produced from multinational projects to foster the relevance and meaningful learning of science and science education (e.g. PROFILES (Bolte, Holbrook, Mamlok-Naaman & Rauch, 2014) or PARSEL (Nielsen et al., 2008)).

However, the important term "relevance" has been interpreted and understood in several ways in the science education community, and thus, there has been a lack of consensus in understanding the meaning of relevance (Stuckey et al., 2013). Likewise, it can be assumed that there is little consensus in developing "relevant" scenarios for CBL. Inspired by the study of Newton (1988), Stuckey et al. (2013) identified three dimensions of relevance in science education—individual, societal, and vocational dimensions—indicating the span of the present-future and the intrinsic-extrinsic range. This model offers a critical insight and several aspects relating to the meaning of relevant science education and its relation to science interest. Notably, in designing scenarios, this model can be considered as categories in measuring, for instance, relevant topics and contents of the scenarios. However, there is insufficient empirical research developing measurement tools of scenarios considering diverse aspects of relevance (e.g. Sormunen, Hartikainen-Ahia & Jäppinen, 2017). Accordingly, this study aims to develop a measurement tool of scenarios and validate it in order to design relevant scenarios for CBL in science education.

## Definitions and Dimensions of Relevance in Science Education

As mentioned, making science classes more relevant to students' daily life has been emphasized in the last few decades because of low levels of students' interest in

science. The concerns relating to science interest have been raised not only for the students who are planning to become future scientists but also for those who are expected to become scientifically literate citizens. Science interest has been related positively to students' involvement in socio-scientific issues that demands scientific understanding both in content and process (Pilot & Bulte, 2006). Particularly, since the 1980s were called "years of crisis in science education," this perspective has constantly been discussed within the science education community. Consequently, various national and international projects have been conducted in an attempt to make science education more relevant. Among them, the Relevance of Science Education (ROSE) project was one of the representative cases (Schreiner & Sjøberg, 2004) that brought the issue of relevant science education to the fore. In their project, the term relevance was defined widely as "meaningful, motivating, interesting, engaging, important, etc." (p. 21), and thus, they investigated a variety of students' motivation in learning science in respect of relevance without any operational definition of the term relevance.

In the last 20 years, issues relating to relevance science education have been discussed rigorously in chemistry education (e.g. King, 2012; Pilot & Bulte, 2006). However, it can be argued that the discussion can be applied to all other science subjects, as chemistry is one of the sciences. Van Aalsvoort (2004a, b) examined a lack of relevance in chemistry education with the lens of logical positivism and activity theory. As the results of the analysis, she suggested four distinguished meanings of relevance as *personal*, *professional*, *social*, and *personal-social relevance* and their implications in teaching chemistry. For instance, in terms of personal relevance, chemical education should make connections between the contents and students' life because "chemical objects, events, and concepts do not have a life of their own, but are connected with products we all use" (Van Aalsvoort, 2004a, p. 1648). In the end, her studies aimed to point out that students should be supported to become responsible citizens by chemical education that embeds the clear contribution of chemistry professions and the social needs with regard to personal-social relevance. In line with Van Aalsvoort, Stuckey et al. (2013) introduced a three-dimensional model of relevance in science education—individual, societal, and vocational dimensions. Specifically, they have specified the meaning of each dimension with two spectra—timeline (present and future) and motivation (intrinsic and extrinsic). As an example, regarding the intrinsic individual dimension, they suggested that science education should consider students' interest (present) and useful skills needed in future for personal life (future). In terms of the extrinsic individual dimension, they demonstrated that good marks in school (present) and acting responsibly in future (future) can be considered to make science lessons more relevant. However, as they indicated, there are many concepts overlapping each other and no clear distinctions between dimensions. Especially, since the term interest can be understood differently and widely in science education, it may be inadequate to put it in one dimension of relevance (individual dimension); indeed, the concept of interest has been related to all other aspects of relevance in their three-dimensional model as well. In other words, in case individual interest is placed among one dimension of relevance, it may conflict the initial idea that relevance in science education should be promoted aiming to increase students' interest (Hulleman & Harackiewicz, 2009). Therefore, these two constructs, relevance and interest, would need to be distinguished in order to have a clear understanding of relationships between relevant science education and students' interest catalyzed by the relevance, as suggested by Kotkas, Holbrook and Rannikmäe (2017).

## The Construct of Science Interest

Interest has been introduced with several theories in educational studies (e.g. Alexander, 2004; Krapp, 2002; Silvia, 2001). However, in contrast to other motivational variables, those theories have related interest to object or specific content such as a particular science content or context, subject, area of knowledge, or activity, and this idea has been dominantly used in the science education community (Potvin & Hasni, 2014).

Regarding the construct of interest, Hidi and Renninger (2006) used three components of interest—*affect*, *knowledge*, and *value*—to introduce the four-phase model of interest development in terms of sequential growth of interest considering the amounts of those three components. Similar to Hidi and Renninger (2006), Krapp (2007) introduced three general characteristics of interest as cognitive aspects (*knowledge*), emotional characteristics (*affect*), and value-related characteristics (*value*). Cognitive aspects or *knowledge* refer to the readiness to acquire new knowledge related to the person's interest since "a person who is interested in a certain subject area is not content with his or her current level of knowledge or abilities in that interest domain" (p. 10). That is, the person with interest may attempt to expand knowledge and learn more about the topic. Emotional characteristics or *affect* refer to positive emotions such as enjoyment connecting with an interest-triggered activity; this positive emotion is often expressed as "I like something or doing something." The value-related characteristics or *value* refer to positive personal evaluation on the object of interest since "a person shows a high subjective esteem for the objects and actions in his or her areas of interest" (p. 11). According to one PISA report (Organization for Economic Cooperation and Development [OECD], 2007), students' personal value of science is clearly distinguished from their social value on science. Moreover, students often see science as an important contributor to societal development; students indicate general support for science needs, but they do not relate science to their lives (OECD, 2007; Salonen, Kärkkäinen & Keinonen, 2018). However, as is known, this personal value of science indicated a strong positive correlation with science performance (OECD, 2007). Interestingly, one of the questions used in PISA measuring the personal value was "science is very relevant to me"; that is, this value-related characteristic of interest may be closely related to the concept of relevant science education.

## Context-Based Learning in Science Education

Given the notion that the use of context in science education is to emphasize students' "learning as relevant to some aspect of their lives" (Gilbert, 2006, p. 960), the movement of context-based learning (CBL) has closely been in line with relevant science education. King (2012) defined CBL specifically in chemistry as "when the 'context' or 'application of the chemistry to a real-world situation' is central to the teaching of the chemistry... that is, when the students require the concepts to understand further the real-world application" (p. 53). Thus, the CBL movement eventually aims to bring science to students' lives (Bennett, Lubben & Hogarth, 2007).

In the 1970s, a large-scale project concerning CBL was started in the Netherlands at the University of Utrecht on the Dutch Physics Curriculum Project (PLON: Project Leerpakket Ontwikkeling Natuurkunde) aimed at connecting physics with students' daily life so as to support students to see science as more attractive and relevant (King,

2012). For this sake, the units of the PLON program started with an orientation concerning social and societal domains of the learner using a basic question that is linked, at last, to relevant science concepts. Followed by the PLON, several other large-scale initiatives have tried to enhance CBL, and Pilot and Bulte (2006) introduced and analyzed five representative CBL-related major projects in different educational systems—Chemistry in Context in the USA, Salters in the UK, Industrial Science in Israel, Chemie im Kontext in Germany, and Chemistry in Practice in the Netherlands. According to their results, all five approaches were developed in order to provide personal and societal relevance to students, and four of them reported that participants enjoyed CBL more than traditional approaches and indicated higher interest in science and more positive opinions about the relevance of science content after CBL experiences. Regarding students' understanding of science concepts, Pilot and Bulte (2006) compared several studies investigating students' conceptual understanding using samples from those five projects and concluded that CBL students indicated a deeper understanding of science concepts than those who participated more traditional approaches. Although some criticism has remained regarding the effects of CBL, there is much evidence reporting positive effects of CBL on students' attitude to and understanding of science (Bennett et al., 2007).

Regarding CBL curricula development, Bennett and Holman (2002) stated that "learner motivation has been the strongest driving force in the development of 'relevant' curriculum materials" since the aim of CBL is to make science more attractive and meaningful to students. For this, the contexts are generally used as the starting point of science lessons to attain and retain students' attention and to develop their scientific ideas contrasting with traditional teaching instructions (Bennett et al., 2007). However, as Gilbert (2006) pointed out, it has been a challenging to relate science content to students' personal relevance in developing context-based curricula. Subsequently, in order to support relevant curricula development, numerous scenarios have been developed and disseminated to the science education community via many large-scale international projects. In the PROFILES (Professional Reflection Oriented Focus on Inquiry-based Learning and Education through Science) project, for instance, a three-stage model approach was introduced consisting of the introduction of relevant socio-scientific scenario as the first stage, inquiry-based learning as the second stage, and students' engagement in a decision-making process relating to the socio-scientific scenario as the last stage (Bolte et al., 2014). During the 5-year project, scenarios relating the content to the contexts were produced and practiced by 21 partner countries, and the scenarios have been publicly open for everyone on the website.

In sum, scenarios in CBL play a pivotal role in connecting school science content with students' daily life so that students can have a positive attitude to learning science. In order to achieve this goal, scenarios should be perceived as relevant and interesting to students, and each construct—relevant and interest—may consist of three components as shown in Fig. 1. However, although much effort has been made to assess the effect of CBL, there is a lack of studies developing measurement tools of the scenario itself considering the relevance and interest of the contents. Since the scenarios have been located in the center of CBL, it is important to evaluate whether the scenario meets the students' needs in terms of relevant and interesting science education. Accordingly, this study aimed to develop and test a measurement tool of scenarios in the lens of the three-dimensional model of relevance and the three characteristics of interest. We also

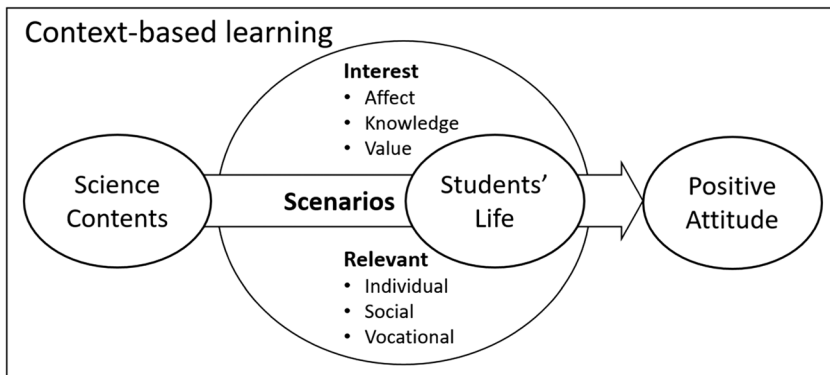


Fig. 1 The role of scenarios in context-based learning

have taken gender and cultural differences into account since these two variables are important in developing and disseminating the tool to multinational studies and environments. This study is a follow-up, extended study of Kotkas et al. (2017) which examined the scenario evaluation instrument with one particular scenario from one particular national sample.

## Method

### Sample and Questionnaire

The sample of this study is from three countries, Estonia, Finland, and the UK, participating the EU project, “Promoting Youth Scientific Career Awareness and Its Attractiveness through Multi-Stakeholder Co-operation” (MultiCO). The MultiCO project aims to make science and science careers more attractive to students; for this, career-based, science-related scenarios concerning diverse global challenges, such as energy, water, waste, food, or health issues, have been developed and introduced in order to increase students’ awareness of science careers and the role of science in society. During the project, five successive interventions using scenarios are being implemented in each country; an instrument, the Scenario Evaluation with Relevance and Interest (SERI) survey for measuring students’ perceptions of career-related scenarios, was developed; before the first intervention, a large-scale survey was carried out to evaluate the scenarios in all participating partners with the instrument, SERI; overall, 25 scenarios were evaluated by students from each country. Students evaluated scenarios as a group or an individual; 574 responses were gathered from 320 girls (or groups of girls) and 253 boys (or groups of boys); 1 respondent did not indicate his or her gender. Most of the participants were aged 13 (see Table 1).

The instrument initially consisted of 28 questions measuring students’ perception of scenarios. According to the result from the small-scale study of Kotkas et al. (2017) using the same instrument, six factors were indicated—*learning value*, *vocational value*, *scenario attributes*, *career awareness*, *social value*, and *like and interest*. However, before a large-scale study, we conducted the rigorous literature review again, built a new conceptual framework, and decided to select 21 questions that were related

**Table 1** Descriptive statistics of the participants

Country	Number of response	Gender ratio
Estonia	177	Girls: 88 Boys: 89
Finland	133	Girls: 57 Boys: 75 Unknown: 1
The UK	264	Girls: 175 Boys: 89
Total	574	Girls: 320 Boys: 253 Unknown: 1

to three dimensions of relevance—individual, societal, and vocational dimensions—and three characteristics of interest—affect, knowledge, and value. In addition, since we found two different aspects of vocational dimensions of relevance in the questionnaire—knowledge gain and future aspiration—we categorized them separately. Each subcategory of relevance consists of three to six questions. Regarding the interest domain, we found four questions reflecting three characteristics of interest and general interest. Finally, we had two categories with eight subcategories as shown in Table 2. While questions relating to the relevance were asked using a four-point Likert scale between totally disagree to totally agree, interest-related questions were asked with a three-point Likert scale considering the nature of the questions and the age of participants (Kotkas et al., 2017). However, since we mostly investigated the correlations of variables in our analyses, these different scales did not affect any of results.

## Data Analysis

We mainly conducted three statistical analyses—exploratory factor analysis (EFA), confirmatory factor analysis (CFA), and multivariate analysis of variance (MANOVA)—in order to achieve our research goals. For the first phase, we conducted EFA and CFA in order to find constructs related to different dimensions of relevance and interest with the 21 questions presented in Table 2. For the factor analyses, the data were split randomly into two halves from each country. The first half ( $n = 286$ ) was used to conduct EFA while the other half ( $n = 288$ ) was used to perform CFA. With the SPSS 23 software, EFA was conducted in order to reduce the items to a smaller number and to investigate the underlying structure of the variables, since EFA is used when the factorial structure of the measuring instruments is unknown (Wang & Wang, 2012). Before conducting EFA, the Kaiser-Meyer-Olkin (KMO) and Bartlett's tests were conducted to check whether the sample was proper for conducting EFA. KMO index above .80 is used to determine whether the sample was appropriate. The principal axis factoring (PAF) extraction method and varimax rotation with a standard eigenvalue greater than 1 criterion were used in EFA. Based on the result of EFA and previous literature, CFA was performed to measure a construct validity using Mplus version 7.4 software (Muthén & Muthén, 1998–2010). Traditional cutoff values were applied in order to assess the goodness of fit: RMSEA (Root Mean Square Error of Approximation) and SRMR (Standardized Root Mean Square Residual) below .08, CFI (comparative fit index) and TLI (Tucker-Lewis Index) above .90 (Wang & Wang, 2012). Then, we checked the internal consistency reliability of each factor with Cronbach's alpha and composite reliability (CR) coefficient to see if they were above .70 in order to assess the

**Table 2** Description of the core variables

Category	Subcategory	Variable name	Item
Relevance	Individual dimension	IND1	This scenario enables me to gain new knowledge about the scenarios' topic.
		IND2	The knowledge I gain from the scenario may be useful in the future.
		IND3	I can put knowledge gained from the scenario into practice, to solve problems.
		IND4	I find this scenario topic important for me personally.
		IND5	I find this scenario topic important to my family.
		<i>IND6</i>	<i>I find this scenario topic important for learning school subjects.</i>
	Societal dimension	SOC1	I find this scenario topic important for appreciating the work of our local community (town, country).
		SOC2	I find this scenario topic important for the whole world.
		SOC3	The scenario presents a scientific problem, which is socially relevant.
	Vocational dimension (knowledge gain)	VKNOW1	From this scenario, I am able to gain new knowledge about possible career(s).
		VKNOW2	This scenario enables me to understand the responsibilities of the persons in the career position indicated
		VNOW3	This scenario enables me to understand the skills that are necessary in this profession.
	Vocational dimension (future aspiration)	VFUT1	I feel my future career may be connected with the topic covered in the scenario.
		VFUT2	I think my future studies at the gymnasium or university level may be connected to the topic covered in the scenario.
		VFUT3	I predict I will need to perform skills, described in the scenario, in my future career.
		VFUT4	I predict I need to perform science-related skills, described in the scenario, in my future career.
		<i>VFUT5</i>	<i>The scenario describes the science community, to which I relate.</i>
Interest	General	INT1	I find this scenario interesting to me.
	Value	INT2	I find the information in this scenario valuable to me.
	Affect	INT3	I liked the scenario.
	Knowledge	INT4	This scenario makes me want to learn more about the topic.

Two variables italics in table, IND6 and VFUT5, were eventually excluded in the measurement after rigorous statistical analyses

internal consistency of each factor. In addition, convergent and discriminant validity were examined based on Fornell and Larcker's (1981) criterion. According to the criterion, the average variance extracted (AVE) value should be higher than .50 to ensure convergent validity, but for a newly developed scale, .45 is viewed reasonable (Netemeyer, Bearden & Sharma, 2003); a squared root value of AVE for each latent



construct should be higher than each latent constructs' highest correlation to confirm discriminant validity.

After assessing validity and reliability, we estimated measurement invariance between genders by means of a multiple group confirmatory factor analysis (MGCFA). Followed by Byrne (2010), we constrained factor loadings and item intercepts across group systematically in a hierarchical order. To begin, configural invariance was examined as a baseline model of the invariance tests. This invariance model tests whether the same number of factors with the same number of items exist across gender with no equality restrictions on other parameters. Thereby, we could confirm the suggested variables in this study explore same constructs across gender. Then, we measured metric invariance by constraining factor loadings across gender and compared the model fit of the metric invariance model to the configural invariance model. If there is no significant difference between the two models, "measures across groups are considered to be on the same scale" (Byrne, 2010, p. 209). Finally, this metric invariance model was compared to a scalar invariance model constraining both factor loadings and item intercepts across groups. No significant differences between the metric and scalar invariance models allow comparing factor means across groups. In these model comparisons, we first checked chi-square ( $\chi^2$ ) differences to see whether the differences were significant between the models. However, since the  $\chi^2$  test has been criticized due to its sensitivity to sample size (Chen, 2007), we also examined changes in CFI ( $\Delta$ CFI), RMSEA ( $\Delta$ RMSEA), and SRMR ( $\Delta$ SRMR) based on the suggestions by Cheung and Rensvold (2002) and Chen (2007). The recommend criteria for invariance were  $\Delta$ CFI  $\leq$  .01,  $\Delta$ RMSEA  $\leq$  .015, and  $\Delta$ SRMR  $\leq$  .015 (metric invariance) or .03 (scalar invariance). After testing and fulfilling the measurement invariance, we conducted the MANOVA in order to find group differences between gender and countries.

## Result

### Exploratory Factor Analysis

The KMO index of EFA was greater than .80 (.92), and Bartlett's test was significant ( $\chi^2 = 4554.50, p < .001$ ) indicating that our data was adequate to conduct EFA; thus, we further investigated other results of EFA. As shown in Table 3, 20 variables loaded greater than .5 remained and formed four factors. However, unlike our assumption, "individual dimension" and "societal dimension" were gathered in one factor; that is, there was a high correlation between these constructs. Also, four variables relating to the three characteristics of interest were indicated as one factor. On the other hand, vocational dimensions were separated into two factors as knowledge gain and future aspiration, as predicted.

In order to examine the correlation of individual dimension and societal dimension that indicated as one factor, we conducted EFA for each country separately to see whether every country indicated a similar pattern or not. For this country-level comparison, we used the full sample size ( $n = 574$ ) because the number of samples from each country became too small to conduct EFA if we only used half of them. Specifically, the number of Finnish samples dropped under 70 while the number of

**Table 3** Exploratory factor analysis for the SERI survey

Category	Subcategory	Variable name	Factor 1	Factor 2	Factor 3	Factor 4
Relevance	Individual dimension	IND1	.62			
		IND2	.71			
		IND3	.64			
		IND4	.62	.55		
		IND5	.71			
	Societal dimension	SOC1	.53			
		SOC2	.57			
		SOC3	.51			
	Vocational dimension (knowledge gain)	VKNOW1		.76		
		VKNOW2		.72		
		VNOW3		.73		
	Vocational dimension (future aspiration)	VFUT1			.82	
		VFUT2			.82	
		VFUT3			.79	
		VFUT4			.71	
VFUT5				.55		
Interest	General	INT1				.85
	Value	INT2				.65
	Affect	INT3				.79
	Knowledge	INT4				.62

variables remained the same as 21; this small sample size may not be adequate for EFA with 21 variables (Everitt, 1975).

According to the EFA comparison, as shown in Table 4, students from Estonia and Finland indicated similar factor loadings on individual and societal dimensions likewise Table 3; only the UK sample presented two distinct factors on these two dimensions. Based on the EFA results, we decided to exclude IND6 and VFUT5 for further analyses.

### Confirmatory Factor Analysis

Based on the results of EFA and previous literature, we constructed CFA model with the samples of Estonia, Finland, and the UK as shown in Fig. 2. CFA was performed with five latent constructs: the individual dimension, societal dimension, two vocational dimensions (knowledge gain and future aspiration), and interest. As shown in Fig. 2, the result presented that the values of factor loading were .54 to .91 exceeding the criterion of .50 (Hair, Black, Babin, Anderson & Tatham, 2006). The measurement model indicated an appropriated model fit (CFI = .931, TLI = .916, RMSEA = .058 (90% C.I. .048 and .068) SRMR = .059). Concerning the convergent validity, the

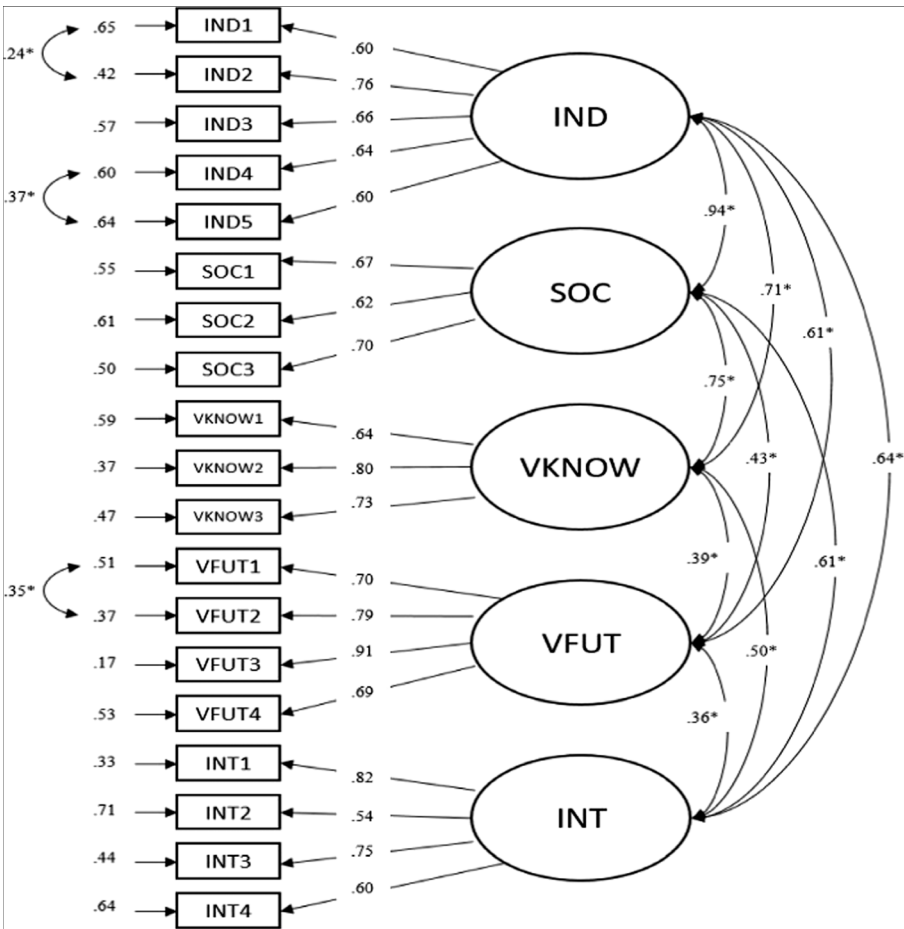
**Table 4** Exploratory factor analysis: comparison table of Estonia, Finland, and the UK

Variable name	Estonia				Finland					UK				
	F1	F2	F3	F4	F1	F2	F3	F4	F5	F1	F2	F3	F4	F5
IND1									.70				.67	
IND2	.54		.56		.60								.77	
IND3	.63				.62								.62	
IND4	.63				.61	.56				.50			.62	
IND5	.78				.71								.57	
IND6	.79				.59									
SOC1			.52		.73									
SOC2	.61				.68									.65
SOC3	.59		.54		.73									.72
VKNOW1			.74					.78			.72			
VKNOW2			.76					.74			.78			
VKNOW3			.65					.61			.76			
VFUT1		.80				.68				.83				
VFUT2		.84				.74				.83				
VFUT3		.81				.83				.83				
VFUT4		.76				.78				.71				
VFUT5		.53								.53				.52
INT1				.81			.61		.55					.82
INT2	.55			.56			.75							.52
INT3				.85					.76					.81
INT4				.57			.80							.67

F, factor

average variance extracted (AVE) results ranged from .43 to .61. Although the AVE values of individual dimension and societal dimension were under .45, they were marginally lower. However, similar to EFA, factors of individual dimension and societal dimension were highly correlated (.94); the correlation value was higher than the squared root values of AVE of individual dimension and societal dimension (Table 5); there can be an issue of discriminant validity between these two constructs in our sample. Regarding the constructs of relevance and interest, they fulfilled the convergent validity and discriminant validity while they also indicated a significant correlation to each other.

With these five factors, internal consistency was examined based on the value of Cronbach's alpha and composite reliability (CR) greater than .70. As presented in Table 5, the Cronbach's alpha and CR reliability coefficient for each scale suggested that all of our scales had satisfactory reliability (> .70) in measuring students' perception on scenarios in terms of relevance and interest.



**Fig. 2** Confirmatory factor analysis for the SERI survey. *Note.*  $\chi^2 = 275.62$  ( $df = 139$ ,  $p < .5$ ), CFI = .931, TLI = .916, RMSEA = .058 (90% C.I. .048 and .068), SRMR = .059. IND, individual dimension; SOC, societal dimension; VKNOW, vocational dimension of knowledge gain; VFUT, vocational dimension of future aspiration; INT, interest. Three residual correlations (IND1-IND2, IND4-IND5, VFUT1-VFUT2) were included on the basis of overlapping item formulations and modification indices identifying potential model fit improvements. \* $p < .05$

In order to confirm two dimensions of relevance and interest, we briefly compared three CFA models, one-factor, five-factor (measurement model), and second-order models as presented in Fig. 3.

According to the result, there was a significant difference between the one-factor model and five-factor measurement model and fit of the one-factor model was unacceptable. On the other hand, while no significant difference was found between five-factor model and second-order model, both indicated proper model fit; second-order model presented slightly better model fit than the five-factor model; thus, overall, the

**Table 5** Validity and reliability of each factor

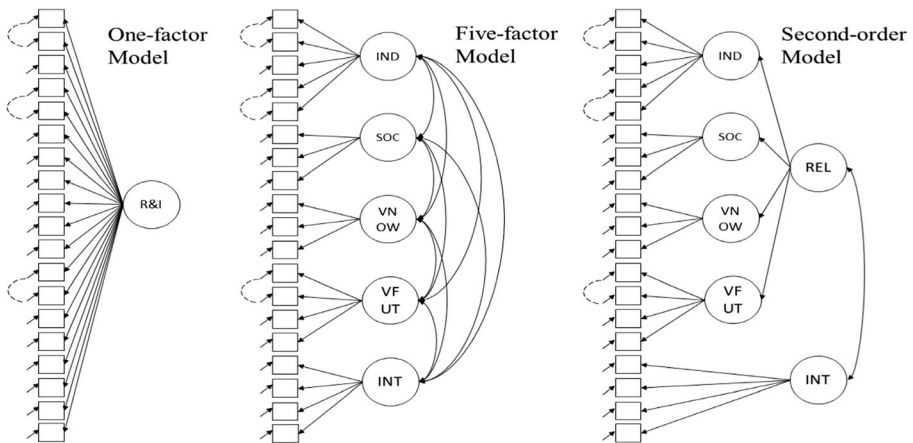
Factor	Cronbach's alpha	CR	AVE	IND	SOC	VKNOW	VFUT	INT
IND	.82	.79	.43	.65				
SOC	.72	.70	.44	.94	.67			
VKNOW	.77	.77	.52	.71	.75	.72		
VFUT	.87	.86	.61	.60	.43	.39	.78	
INT	.79	.78	.47	.64	.61	.50	.36	.69

*Italic numbers show the each squared root value of AVE. IND, individual dimension; SOC, societal dimension; VKNOW, vocational dimension of knowledge gain; VFUT, vocational dimension of future aspiration; INT, interest*

result confirmed two separate dimensions of relevance and interest in evaluating scenarios (see Table 6).

**Measurement Invariance Across Gender**

As shown in Table 7, the configural model indicated an acceptable fit for all the goodness-of-fit values providing adequate support for configural invariance of the suggested measurement. Regarding the model comparison between configural and metric and between metric and scalar,  $\chi^2$  tests indicated statistical differences for both comparisons. However, most of the goodness-of-fit values remained the same; that is,  $\Delta CFI \leq .01$ ,  $\Delta RMSEA \leq .015$ , and  $\Delta SRMR \leq .015$  Therefore, this finding supported that none of these constraints resulted in a model fit decrease, and thus, the measurement invariance of the SERI survey across gender was fulfilled.



**Fig. 3** Three competing CFA models

**Table 6** CFA model comparison

Model	$\chi^2$ ( <i>df</i> )	CFI	TLI	RMSEA (90% C.I.)	SRMR	Reference model	SB- $\Delta\chi^2$ ( $\Delta df$ )	$\Delta CFI$
A. One-factor	665.36* (149)	.740	.702	.110 (.101, .118)	.090			
B. Five-factor	275.62* (139)	.931	.916	.058 (.048, .068)	.059	A	320.6 (10)*	.191
C. Second-order	271.26* (144)	.936	.924	.055 (.045, .065)	.061	B	-2.6 (5)	.005

SB- $\Delta\chi^2$ , Satorra-Bentler corrected  $\chi^2$  value (Satorra & Bentler, 2010); \* $p < .05$

## Multivariate Analysis of Variance

Finally, we investigated group differences via MANOVA by examining the mean differences in relevance and interest towards scenarios by gender and countries (Estonia, Finland, and the UK) and the results were derived based on estimated marginal means. As shown in Table 8, regarding 2.5 as a central value, students indicated that presented scenarios were moderately relevant individually, socially, and vocationally. However, students responded that science careers introduced in the scenarios were not likely to be related or relevant to their future career aspiration.

Table 9 presents group differences in scenario evaluations concerning relevance and interest. There were no significant gender differences with regard to the relevance of scenarios while significant gender differences were found in accordance with interest in scenarios. On the other hand, statistical group differences were found at the country level among three countries, Estonia, Finland, and the UK, regarding both relevance and interest in scenarios, except the vocational dimension in future aspiration ( $p = .055$ ). Thus, we conducted an independent  $t$  test and a post hoc test to investigate where these gender and country-level differences were derived from.

In order to check the gender gap with regard to interest in scenarios, independent  $t$  test was conducted for each country. As shown in Table 10, while there were no gender differences in Finland and the UK, Estonian students indicated a significant gender difference in interest in scenarios (.78,  $p < .05$ ) that boys' interest was very low compared to girls and students of other countries.

Table 11 presents the mean differences between Estonia, Finland, and the UK. Regarding the individual dimension, Estonia indicated a significant difference with Finland and the UK; regarding the societal dimension, Estonia indicated a significant difference with the UK; regarding the vocational dimension of knowledge gain, all three countries indicated differences with each other; regarding the vocational

**Table 7** Measurement invariance: gender

Model	$\chi^2$ ( <i>df</i> )	CFI	RMSEA	SRMR	$\Delta\chi^2$ ( $\Delta df$ )	$\Delta CFI$	$\Delta RMSEA$	$\Delta SRMR$
Configural	500.42* (278)	.895	.075	.070				
Metric	530.38* (292)	.888	.075	.081	30.77 (14)*	.007	.000	.011
Scalar	548.64* (306)	.886	.074	.081	48.23 (28)*	.002	.001	.000

\* $p < .05$

**Table 8** Grand mean of variables

Category	Subcategory	Mean (SD)	95% confidence interval	
			Lower bound	Upper bound
Relevance	Individual dimension	2.42 (.03)	2.37	2.47
	Societal dimension	2.58 (.03)	2.53	2.64
	Vocational dimension (knowledge gain)	2.69 (.03)	2.62	2.74
	Vocational dimension (future aspiration)	2.07 (.03)	2.01	2.13
Interest		2.33 (.04)	2.25	2.41

dimension in future aspiration, there were no significant differences between countries; lastly, with regard to interest in scenarios, Estonia again indicated a significant difference with Finland and the UK probably due to boys as shown in Table 10.

Overall, Estonian students presented some differences in regard to their perspectives on relevance and interest in the introduced scenarios while Finnish and the UK students presented similarities. In addition, as shown in Fig. 4, these differences resulted from Estonian students' negative responses on the questionnaire.

Thus, we further conducted an independent *t* test with the Estonian sample for all five factors to measure the gender differences. The mean comparisons using *t* test with Bonferroni correction indicated that boys presented such negative responses to all five aspects that overall value of Estonian sample indicated a significant difference with Finnish and the UK samples (see Table 12). That is, Estonian students' gender differences in their perspectives on scenarios caused country-level differences with Finland and the UK eventually.

**Table 9** Between-group differences

Group	Category	Subcategory	Mean square	<i>F</i>
Gender	Relevance	Individual dimension	.15	.42
		Societal dimension	.43	.95
		Vocational dimension (knowledge gain)	.17	.43
		Vocational dimension (future aspiration)	.88	1.91
	Interest		7.94***	10.70
Country	Relevance	Individual dimension	2.60**	7.33
		Societal dimension	2.98**	6.83
		Vocational dimension (knowledge gain)	4.61***	11.43
		Vocational dimension (future aspiration)	1.35 <sup><i>p</i> = .055</sup>	2.92
	Interest		6.83*	9.20

\**p* < .05, \*\**p* < .005, \*\*\**p* < .001

**Table 10** Independent *t* test between genders within countries concerning interest in scenarios

Country	<i>t</i> ( <i>dF</i> )	Mean	Mean difference (S.E)
Estonia	5.92 (175)	Girls: 2.50 Boys: 1.72	.78*** (.13)
Finland	1.23 (122)	Girls: 2.64 Boys: 2.46	.18 (.14)
UK	.82 (262)	Girls: 2.40 Boys: 2.30	.10 (.12)

\*\*\**p* < .001

## Discussion

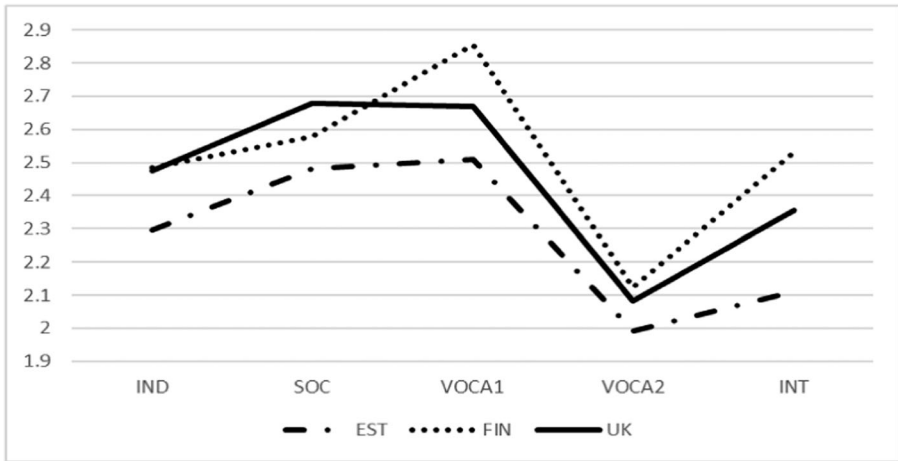
While CBL has been recommended and expanded in science education (Gilbert, 2006; King, 2012; Pilot & Bulte, 2006) and scenarios have been developed and introduced as the starting point of science lessons to relate science content to daily contexts in CBL approaches (Bolte et al., 2014; Bennett et al., 2007; Nielsen et al., 2008), there is a lack of resources to measure good scenarios that have been placed at the center of CBL in terms of relevance and interest. Therefore, this study aimed to develop and validate an instrument to examine students' perspectives on science career-related scenarios reflecting three dimensions of relevance and three characteristics of interest. According to the previous literature, the survey, and the data analyses, we have found four relevance-related factors—individual, societal, vocational knowledge gain,

**Table 11** Post hoc test between countries

Category	Subcategory	Country (I)	Country (J)	I-J (S.E)
Relevance	Individual dimension	Estonia	Finland	-.19* (.07)
		Estonia	UK	-.18** (.06)
		Finland	UK	.01 (.07)
	Societal dimension	Estonia	Finland	-.11 (.08)
		Estonia	UK	-.20** (.07)
		Finland	UK	-.09 (.07)
	Vocational dimension (knowledge gain)	Estonia	Finland	-.35*** (.08)
		Estonia	UK	-.16* (.06)
		Finland	UK	.18* (.07)
	Vocational dimension (future aspiration)	Estonia	Finland	-.13 (.11)
		Estonia	UK	-.09 (.18)
		Finland	UK	-.04 (.61)
Interest		Estonia	Finland	-.42*** (.10)
		Estonia	UK	-.25* (.09)
		Finland	UK	.17 (.10)

\**p* < .05, \*\**p* < .005, \*\*\**p* < .001





**Fig. 4** Comparison of three countries in students’ perspectives of scenarios. Note. IND, individual dimension; SOC, societal dimension; VKNOW, vocational dimension of knowledge gain; VFUT, vocational dimension of future aspiration; INT, interest

and vocational future aspiration—and one interest-related factor. This newly invented instrument has been investigated by three different country samples—Estonia, Finland, and the UK—in order to measure gender and country differences. According to the results, this instrument has good construct validity and reliabilities. In addition, the constructs of relevance and interest were perceived differently by students although these two constructs have significant correlations with each other. Also, gender differences were found in the Estonian sample.

The instrument, the scenario evaluation with relevance and interest (SERI) survey for measuring students’ perceptions on career-related scenarios, was designed based on three dimensions of relevance by Stuckey et al. (2013) and three characteristics of interest by Krapp (2007) and Hidi and Renninger (2006). According to the rigorous

**Table 12** Independent *t* test between genders: Estonian sample

Category	Subcategory	<i>t</i> ( <i>df</i> )	Mean	Mean difference (S.E)
Relevance	Individual dimension	3.35 (175)	Girls: 2.46 Boys: 2.12	.34** (.10)
	Societal dimension	2.75 (175)	Girls: 2.63 Boys: 2.31	.32* (.11)
	Vocational dimension (knowledge gain)	3.41 (175)	Girls: 2.69 Boys: 2.31	.38* (.11)
	Vocational dimension (future aspiration)	1.76 (174)	Girls: 2.09 Boys: 1.90	.18 <sup><i>p</i> = .08</sup> (.10)
Interest		5.92 (175)	Girls: 2.50 Boys: 1.72	.78*** (.13)

\**p* < .05, \*\**p* < .005, \*\*\**p* < .001

statistical analyses based on the samples from three countries, this instrument is well qualified concerning reliability and validity. However, we found one issue of discriminant validity between two constructs, individual dimension and societal dimension of relevance, from Estonian and Finnish samples. This issue with discriminant validity in these two countries may result from a small number of participants. There are no specific guidelines regarding minimum sample size for conducting factor analyses; however, it has been recommended to have a ratio of 10 times the number of variables (Everitt, 1975); in this study, we used 21 variables; that is, each country needed a minimum of 210 samples according to the recommendation, but Estonia and Finland had 177 and 133 respectively while the UK included 264 responses. Interestingly, the UK indicated different factor loadings with the other two countries, and the individual and societal dimensions were grouped in different factors. However, the recommended sample size is not only about the ratio but also, “is related to the number of variables, the number of factors, the number of variables per factor, and the size of the communalities” (Mundfrom, Shaw & Ke, 2005, pp. 160–161) and still it is inconclusive. Another possible explanation for this issue can be attributed to the students’ age group since “the individual dimension might be more important for younger children, but this importance will shift towards societal relevance as the child grows and matures” (Stuckey et al., 2013, p. 24). Students participating in this survey were mostly 13-year-olds and this year of the age group has been deemed as a transition period to adolescence. Therefore, there may be a chance to see these individual dimension and societal dimension of relevance as similarly important from the participants. For further research, subsequently, it is recommended to investigate different age groups, such as students in upper secondary school, with the same instrument to see whether the pattern of factor loadings is different.

Concerning the constructs of relevance and interest, in contrast to Stuckey et al. (2013) suggesting interest as a part of the individual dimension of relevance, this study found that students clearly distinguished relevance and interest as different constructs in measuring the scenarios, and these two constructs indicated a moderated but statistically significant relationship. This result is in accord with recent studies indicating that when students were encouraged to relate science courses to their lives, their interest in science was increased (Bennett & Holman, 2002; Hulleman & Harackiewicz, 2009). As Kotkas et al. (2017) argued, although the concepts of relevance and interest are closely related and overlap each other, it is not justified to consider these constructs as synonyms or interest as a part of the relevance. Rather, our result suggests that it would be beneficial to measure students’ perspective on scenarios in terms of interest and relevance separately and to find their relationship if they are highly correlated, since it may be possible that the scenarios are relevant but not interesting and vice versa.

Regarding the group differences, we found country-level differences in students’ perspectives on relevant and interesting scenarios resulted from low scores by Estonian boys. While the UK and Finnish students and Estonian girls indicated moderate scores in evaluating scenarios, Estonian boys presented a very negative trend in their responses and all those responses were significantly lower than other participants statistically. According to the previous study of Teppo and Rannikmäe (2003) from a different project, Estonian students indicated gender differences in their preference of scenarios that girls preferred scenarios that dealt with health and outlook while boys have more interested in scenarios linked to economic and environmental problems.

Since the aim of this study was not to investigate students' preference for different scenarios, we did not examine gender differences in each developed scenario. However, it must be valuable to evaluate how each scenario with different topics affects students' perceptions differently in terms of relevance and interest.

Relating to the MultiCO project, the scenarios were perceived as moderately relevant and interesting to participating students. In particular, the MultiCO scenarios were highly valued in terms of their societal relevance and vocational knowledge gain. On the other hand, with respect to the vocational future aspiration, students indicated that these scenarios were not relevant to their future careers. Although, in general, less than 25% of students expect to work in an occupation requiring science-related skills (OECD, 2016), given that the project aims to increase students' interest in science careers, this result may be problematic to achieve the goal of the MultiCO project with the scenarios. Considering the high correlation of vocational relevance with individual dimension, it is recommended to modify the scenarios in a way to increase the individual relevance score of students as well as the vocational dimension of future aspiration guided by the questions suggested in the SERI survey.

## Conclusion and Limitation

In creating, developing, and modifying scenarios for CBL, educators should evaluate the scenarios through a proper measurement tool. This newly designed instrument, SERI, may be valuable for educators such as teachers or scenario developers as a tool to measure students' perceptions of scenarios in implementing CBL in science education. Knowing students' attitudes on scenarios may support educators to modify their instructions so as to draw students' attention to science more and to make their class successful. In addition, since relevance and interest are the most important indicators in designing scenarios and preparation of CBL, this tool will help teachers consider these constructs continuously in their science lesson. Notably, this tool can be used in multinational projects developing and using scenarios for the projects. Although there are attempts to design worldwide scenarios in multinational projects like MultiCO or PROFILES, since students are placed in different social, cultural, and environmental situations, there may be different interest and understanding towards scenarios in different countries and cultures. Thus, it is important to measure whether the scenarios indicate gender and cultural differences beforehand as we have done with SERI instrument. Indeed, results from different cultures may require to revise this tool. So we recommend conducting research using this tool with different samples from other cultural backgrounds in order to generalize or modify the results.

This study is not without limitations. One of the limitations is this study did not measure whether the highly rated scenarios ultimately lead to students' higher achievements or interest in science since we only measured students' interest in science-related scenarios, not science interest. However, as is known, the effects of scenarios will be varied based on how they will be presented by instructors. Also, as the previous studies indicated it could be assumed that if students see the scenarios as more relevant to their lives, their interest in science will be increased, and the increased interest in science might affect their achievements as well (Hulleman & Harackiewicz, 2009; OECD, 2007). In addition, as we discussed, this tool indicated the issue of discriminant validity

between two factors, individual and societal dimensions of relevance from two participating countries. Therefore, it again is recommended to research this tool with different country samples and age groups. In addition, regarding the sampling collection, although we used multinational samples, they were from one specific project. Therefore, in order to generalize the result, unbiased large-scale sampling as has been done by PISA or TIMSS should be conducted.

**Acknowledgments** Open access funding provided by University of Eastern Finland (UEF) including Kuopio University Hospital. We thank to all MultiCO project team members for their contribution to translating the questionnaire and collecting the data. This project has received funding from the *European Union's Horizon 2020 research and innovation program* under grant agreement no. 665100.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Alexander, P. A. (2004). A model of domain learning: Reinterpreting expertise as a multidimensional, multistage process. In D. Y. Dai & R. J. Sternberg (Eds.), *Motivation, emotion, and cognition: Integrative perspectives on intellectual functioning and development* (pp. 273–298). Mahwah, NJ: Lawrence Erlbaum Associates, Inc..
- Bennett, J., & Holman, J. (2002). Context-based approaches to the teaching of chemistry: What are they and what are their effects? In J. K. Gilbert, O. de Jong, R. Justi, D. F. Treagust, & J. H. van Driel (Eds.), *Chemical education: Towards research-based practice* (pp. 165–184). The Netherlands: Kluwer Academic Publishers.
- Bennett, J., Lubben, F., & Hogarth, S. (2007). Bringing science to life: A synthesis of the research evidence on the effects of context-based and STS approaches to science teaching. *Science Education*, *91*(3), 347–370.
- Bolte, C., Holbrook, J., Mamluk-Naaman, R., & Rauch, F. (2014). *Science teachers' continuous professional development in Europe. Case Studies from the PROFILES Project*. Berlin: Freie Universität Berlin Germany/Klagenfurt: Alpen-Adria-Universität Klagenfurt Austria.
- Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Abingdon, England: Routledge.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464–504.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233–255.
- Everitt, B. S. (1975). Multivariate analysis: The need for data, and other problems. *British Journal of Psychiatry*, *126*(3), 237–240.
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39–50.
- Fouad, N. A., Hackett, G., Smith, P. L., Kantamneni, N., Fitzpatrick, M., Haag, S., & Spencer, D. (2010). Barriers and supports for continuing in mathematics and science: Gender and educational level differences. *Journal of Vocational Behavior*, *77*(3), 361–373.
- Gilbert, J. K. (2006). On the nature of 'context' in chemical education. *International Journal of Science Education*, *28*(9), 957–976.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (6th ed.). Upper Saddle River, NJ: Prentice Hall.
- Hidi, S., & Renninger, A. (2006). The four-phase model of interest development. *Educational Psychologist*, *41*(2), 111–127. [https://doi.org/10.1207/s15326985ep4102\\_4](https://doi.org/10.1207/s15326985ep4102_4).
- Holbrook, J. (2008). Introduction to the special issue of science education international devoted to PARSEL. *Science Education International*, *19*(3), 257–266.

- Hulleman, C. H., & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. *Science*, *326*(5958), 1410–1412.
- Kang, J., & Keinonen, T. (2017). The effect of inquiry-based learning experiences on adolescents' science-related career aspiration in the Finnish context. *International Journal of Science Education*, *39*(12), 1669–1689.
- King, D. (2012). New perspectives on context-based chemistry education: Using a dialectical sociocultural approach to view teaching and learning. *Studies in Science Education*, *48*(1), 51–87.
- Kotkas, T., Holbrook, J., & Rannikmäe, M. (2017). A theory-based instrument to evaluate motivational triggers perceived by students in stem career-related scenarios. *Journal of Baltic Science Education*, *16*(6), 836–854.
- Krapp, A. (2002). An educational-psychological theory of interest and its relation to self-determination theory. In E. Deci & R. Ryan (Eds.), *The handbook of self-determination research* (pp. 405–427). Rochester, NY: University of Rochester Press.
- Krapp, A. (2007). An educational–psychological conceptualisation of interest. *International Journal for Educational and Vocational Guidance*, *7*(1), 5–21.
- Lubben, F., Bennett, J., Hogarth, S., & Robinson, A. (2004). A systematic review of the effects of context-based and science–technology–society (STS) approaches in the teaching of secondary science on boys and girls, and on lower ability students. In *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education. Retrieved August 24, 2006, from [eppi.ioc.ac.uk/EPPI](http://eppi.ioc.ac.uk/EPPI).
- Mundfrom, D. J., Shaw, D. G., & Ke, T. L. (2005). Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing*, *5*(2), 159–168.
- Muthén, L. K., & Muthén, B. O. (1998–2010). *Mplus user's guide*. 6th ed. Los Angeles, CA: Author.
- Newton, D. P. (1988). Relevance and science education. *Educational Philosophy and Theory*, *20*, 7–12.
- Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling procedures: Issues and applications*. Thousand Oaks, CA: Sage Publications.
- Nielsen, J., Lindner, M., Bolte, C., Galvão Couto, C., Gräber, W., Holbrook, J., Mamlok, R., ... Tsarपालis, G. (2008). Popularity and Relevance of Science Education and Scientific Literacy – The PARSEL Project in Europe. In A. Macdonald (Ed.), *Planning science instruction: From insight to learning to pedagogical practices. Proceedings of the 9th Nordic Research Symposium on Science Education, 11th–15th June 2008, Reykjavik, Iceland* (pp. 193–194). Reykjavik, Iceland: Science Education Research Group, School of Education, University of Iceland.
- Organization for Economic Co-operation and Development (2007). *PISA 2006: Science competencies for tomorrow's world volume 1: Analysis*. Paris, France: OECD.
- Organization for Economic Co-operation and Development (2016). *PISA 2015 results (volume I): Excellence and equity in education*. Paris, France: OECD.
- Pilot, A., & Bulte, A. M. (2006). The use of “contexts” as a challenge for the chemistry curriculum: Its successes and the need for further development and understanding. *International Journal of Science Education*, *28*(9), 1087–1112.
- Potvin, P., & Hasni, A. (2014). Interest, motivation and attitude towards science and technology at K-12 levels: A systematic review of 12 years of educational research. *Studies in Science Education*, *50*(1), 85–129.
- Salonen, A., Kärkkäinen, S., & Keinonen, T. (2018). Career-related instruction promoting students' career awareness and interest towards science learning. *Chemistry Education Research and Practice*, *19*(2), 474–483.
- Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika*, *75*(2), 243–248.
- Schoon, I. (2001). Teenage job aspirations and career attainment in adulthood: A 17-year follow-up study of teenagers who aspired to become scientists, health professionals, or engineers. *International Journal of Behavioral Development*, *25*(2), 124–132.
- Schreiner, C., & Sjøberg, S. (2004). *Relevance of science education: Sowing the seeds of ROSE*. Oslo, Norway: Acta Didactica.
- Silvia, P. J. (2001). Interest and interests: The psychology of constructive capriciousness. *Review of General Psychology*, *5*(3), 270–290.
- Simon, S., & Osborne, J. (2010). Students' attitudes to science. In J. Osborne & J. Dillon (Eds.), *Good practice in science teaching: What research has to say* (2nd ed., pp. 238–258). Maidenhead, England: Open University Press.
- Sormunen, K., Hartikainen-Ahia, A., & Jäppinen, I. (2017). Quality of SSI scenarios designed by science teachers. In K. Hahl, K. Juuti, J. Lampiselkä, A. Uitto, & J. Lavonen (Eds.), *Cognitive and affective aspects in science education research* (pp. 103–116). Cham, Switzerland: Springer.

- Stuckey, M., Hofstein, A., Mamlok-Naaman, R., & Eilks, I. (2013). The meaning of 'relevance' in science education and its implications for the science curriculum. *Studies in Science Education*, 49(1), 1–34.
- Taconis, R., & den Brok, P. (Eds.). (2016). *Teachers creating context-based learning environments in science*. Cham, Switzerland: Springer.
- Tai, R. H., Liu, C. Q., Maltese, A. V., & Fan, X. (2006). Career choice: Planning early for careers in science. *Science*, 312(5777), 1143–1144.
- Teppo, M., & Rannikmäe, M. (2003). Increasing the relevance of science education – student preferences for different types of teaching scenarios. *Journal of Baltic Science Education*, 2(2), 49–61.
- Van Aalsvoort, J. (2004a). Activity theory as a tool to address the problem of chemistry's lack of relevance in secondary school chemical education. *International Journal of Science Education*, 26(13), 1635–1651.
- Van Aalsvoort, J. (2004b). Logical positivism as a tool to analyse the problem of chemistry's lack of relevance in secondary school chemical education. *International Journal of Science Education*, 26(9), 1151–1168.
- Wang, J., & Wang, X. (2012). *Structural equation modeling: Applications using Mplus*. Hoboken, NJ: Wiley.