International Journal of Science and Mathematics Education (2023) 21:1725–1747
https://doi.org/10.1007/s10763-022-10325-y

Gendered Science Practice at Secondary School and its Effects on Science Motivations

Jingoo Kang[1]

Received: 23 February 2022 / Accepted: 20 September 2022 / Published online: 15 October 2022

Abstract

This study explores the effects of gendered inquiry-based learning (IBL) practices on other science learning experiences and motivations including STEM (Science, Technology, Engineering, and Mathematics) career aspirations. For this, PISA 2015 data representing the Finnish 15-year-old student population has been selected and analyzed by multiple group structural equation modeling. According to the results, while girls and boys perceived a similar frequency of conducting IBL at school as a student, boys were more aware that students were allowed to conduct high autonomous IBL activities such as designing their own experiments. Also, boys expressed that they got more feedback from their teachers, and it was moderately related to the IBL engagement. Moreover, the result showed that while boys perceived more IBL experiences, the efficacy of the IBL activities was higher for girls since the relationship between IBL and science motivations was stronger for girls compared to boys. Lastly, this study found that a teacher was more prone to undervalue girls' science performance than boys during IBL. Based on the findings, this study argues that school IBL practices could be a good source for examining gender equity in science education. Also, the findings recommend gender-inclusive science practices at school in order to increase girls' interest in STEM careers and to enhance the effects of IBL on students' science achievement.

Keywords Gender · Inquiry-based learning · Interaction · Motivation · STEM careers

Introduction

In the contemporary world, science and technology have become a foundation of overall industries; thus, careers requiring STEM (Science, Technology, Engineering, and Mathematics) skills are becoming more abundant in the job markets

✉ Jingoo Kang
jingoo.kang@uef.fi

[1] School of Applied Educational Science and Teacher Education, University of Eastern Finland, Yliopistokatu 2, 80100 Joensuu, Finland

(National Science Board, 2015). Given that the single most important factor affecting individuals' career aspirations is expected later labor market returns (Moorhouse, 2017), it was expected to have enough students pursuing STEM studies at college as the market grows (Sithole et al., 2017). However, ironically, in the last two decades, students, especially in secondary education, have indicated a negative trend in their interest in science and STEM careers especially in many western developed countries (Kang et al., 2019a; Organization for Economic Cooperation and Development [OECD], 2016a), and consequently, there is a lack of college students who pursue STEM majors such as physics, engineering, or computer science requiring advanced mathematic skills (Giffi et al., 2015). Interestingly, these are the majors that have been dominated by males in higher education for the past 20 years (Stoet & Geary, 2018); thus, the deficiency of the STEM workforce might be related to and addressed with gender perspectives to properly understand, intervene, and improve the situation (Holmes et al., 2018).

Up to date, much research has been conducted to reveal influencing factors of students' engagement in STEM fields and found that among many factors, students' experiences in inquiry-based learning (IBL) may play an important role in fostering science aspirations for young students (Cairns & Areepattamannil, 2019; Kang & Keinonen, 2017) in addition to interest, self-efficacy, and outcome expectations (Jeffries et al., 2020; Kang et al., 2021a; Potvin & Hasni, 2014). However, these factors are not working individually, but collaboratively in developing STEM aspirations as Social Cognitive Career Theory (SCCT by Lent et al., 1994) describes. However, recent studies keep indicating that girls possessed a lower interest in science and STEM careers as well as lower self-efficacy compared to boys (Kang, 2022; OECD, 2016a, 2019). In addition, girls got less support and encouragement during science lessons from their teachers (Archer et al., 2020). Besides, girls' and boys' experiences in IBL were different as boys played a more active role while girls became passive during the experiments (Wieselmann et al., 2020). Therefore, it is important to get a deeper understanding of how IBL are practiced at secondary school concerning gender disparities and to what extent the IBL affects students' STEM aspirations differently by gender as well as interest, self-efficacy, and teacher support.

Accordingly, this study aimed to explore the effects of IBL on students' aspirations toward STEM and what roles gender plays in this regard using PISA 2015 datasets. Specifically, based on the SCCT model, the relationships between IBL, self-efficacy, outcome expectations, interest, and career aspirations were examined as shown in Fig. 1. In addition, the effects of teachers' individual feedback and general support during science lessons were investigated as a proxy of the teacher-student interaction. To this end, this study tried to answer the following research questions:

RQ1. To what extent do girls and boys perceive differently in their IBL experiences and teacher support during science lessons? Based on the previous research, it is hypothesized that while the girls and boys indicated similar perceptions of students' participation in IBL activities, boys may indicate more than girls that students were asked to involve in certain types of IBL activities, such as

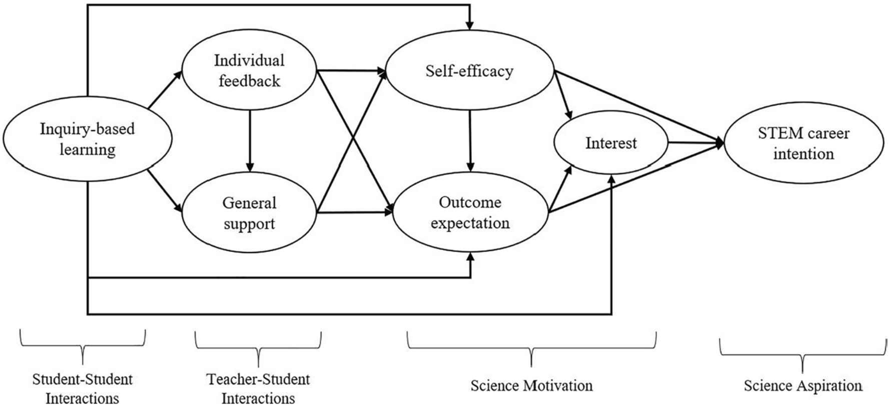

Fig. 1 Hypothesized statistical model measuring the effects of IBL on STEM aspirations

a discussion or testing their own idea, than girls. Also, boys may perceive more than girls that students experienced interactions with their teachers during science lessons in the form of support or feedback.

RQ2. To what extent do the relationships between IBL experiences, science motivations, and aspirations differ by gender? It is hypothesized that the impact of IBL on STEM aspirations may be mediated by interest, self-efficacy, and outcome expectations. However, since girls may be less active than boys and they may indicate fewer interactions with the teacher during IBL, the relationships between IBL and other factors are different between genders. Specifically, the effect of IBL on STEM aspirations may be stronger for boys than girls, since teacher-student interactions may enhance the relationship.

Effect of IBL on Interest, Self-efficacy, Outcome Expectation, and STEM Aspirations

For the last decades, IBL has been placed in the heart of science education since it resembles authentic works of scientists and indicates positive effects on science learning both on knowledge and process (Kang, 2022; Şimşek & Kabapınar, 2010). In addition, several studies pointed out that students' participation in IBL increased their interest (Kang & Keinonen, 2018; Oliver et al., 2021), self-efficacy (Çalık, 2013; Kang & Keinonen, 2017), and aspirations toward STEM studies and careers (Cairns & Areepattamannil, 2019). Accordingly, IBL has been accepted as a core curriculum in teaching science subjects and become a common practice in science education internationally due to its positive effects on learning outcomes (Kang, 2022).

The associations of interest, self-efficacy, and STEM aspirations with IBL were often explained by the SCCT model (Kang & Keinonen, 2017; Taskinen et al., 2013). SCCT introduced by Lent and his colleagues is a theoretical framework presenting the dynamic interaction of personal and environmental factors in explaining an individual's career aspirations or behaviors (Lent et al., 1994).

According to the theory, personal learning experiences can foster self-efficacy and outcome expectations as well as interest as shown in Fig. 1. For instance, participation in science clubs or extracurricular science activities can influence students' interest in STEM studies and careers since those learning experiences can foster students' confidence in science and expectations of outcomes by engaging in the behavior (Taskinen et al., 2013). However, the SCCT framework points out that both learning experiences and career goals can be moderated by person inputs such as gender, race, or predispositions (Almukhambetova et al., 2021). Especially, Miller et al. (2018) reported that girls' science interest starts to decrease when the science lessons at school become more structured. Thus, it is assumed that gender plays a certain role in school science practice, and this gendered science practice contributes to students' STEM aspirations. Thus, the SCCT offers a proper framework that we can explore the effect of school science practices on students' motivations and what role gender plays in this regard.

Recent international large-scale studies (ILSA), however, bring doubt on the use of IBL due to its significant negative association with science achievement in most of the participating countries (OECD, 2016b). After the publication of the results, severe debates have arisen over whether we should stop using IBL to climb up to the PISA ranking (e.g. Sjøberg, 2017). However, similar results were reported in the UK that not only PISA scores but also national standardized test scores in science were negatively associated with IBL participation (Jerrim et al., 2022). Concerning this incongruity between the current large-scale evidence of the negative effects of IBL from ILSA and the traditional belief on the positive effects of IBL, researchers have tried to give an explanation by segregating the effects of different levels of IBL on science achievement. That is, while a basic level of IBL indicated a positive effect on knowledge gain, more complex levels of IBL could be associated negatively with science performance (Aditomo & Klieme, 2020; Kang & Keinonen, 2018). This claim was often based on the study by Kirschner et al. (2006) that argues that the high level of IBL might increase the cognitive burden that impedes student learning abilities.

Although separation of the levels of IBL based on the complexity explained a part of the incongruity, however, what is still widely unknown is the gender effect on the relationship between IBL and achievement. That is, as described by the SCCT model, girls and boys may have different IBL experiences that may moderate the effects of IBL on achievement either positive or negative. Up to date, most of the findings concerning gendered IBL participation and its effects are based on classroom observations using qualitative methods (e.g. Brotman & Moore, 2008; Wieselmann et al., 2020). One study by Lee and Burkam (1996), a quantitative study to investigate whether 8th graders' laboratory experiences were different by gender, found that both of them equally participated in regular laboratory experiments, while only girls' lab experiences were positively associated with their achievement, but not boys'. However, their study only asked students if they received science lab at least once a week as a dummy variable, but did not ask about any other specific type of activities in detail such as designing the experiment, making conclusions, or discussion during IBL that may be biased by gender. Thus, it is recommended to conduct more quantitative studies using more detailed scales measuring a different type of IBL activities to investigate gender equity in IBL engagement.

Impact of Gendered IBL Practices

As stated, according to the SCCT framework, students' IBL experiences can differ by gender. One of the examples of gendered science practices at school is presented by Wieselmann et al. (2020) explaining how small group science activities can be easily dominated by boys. Wieselmann and her colleagues observed four fifth graders, two girls and two boys, participating in small group activities focusing on integrated STEM curricula. During 14 days of unit implementation, they found that boys mainly played a leadership role while girls participated as a supporter. Specifically, girls were more observing and recording the process, while boys were more directing, initiating, and reasoning; boys took physical control of the materials twice more than girls and control of conversations; many times, girls' opinions were unacknowledged while boys were more respected and accepted during the student–student interactions. Therefore, we can assume that this kind of male-dominated science practice may happen regularly at school that affects girls' perception of science as male-friendly which may deter girls' aspiration in science studies and careers.

In addition to these student–student interactions, teacher-student interactions can influence students' perceptions of science because the teacher-student interaction is found to be a strong predictor of students' attitudes toward science and their learning goals in science (Hattie, 2009). The teacher-student interactions at school may happen in two forms: (1) individual feedback after carrying out a certain type of activity such as IBL for improving student learning outcomes (Lipko-Speed et al., 2014) and (2) general support for the whole classroom to create a classroom environment more conducive to learning (Ma & Willms, 2004). The type of individual feedback was found to be praise, approval, rewards, or corrective (Hattie & Timperley, 2007), but it provides information about the task to guide individual students in a proper way (Deci et al., 1999) or to modify their behaviors for better learning outcomes (OECD, 2016b). A recent meta-analysis concerning feedback indicates that it has a higher impact on cognitive outcomes than motivational or behavioral outcomes (Wisniewski et al., 2020). In addition to individual feedback, a teacher could create a supportive climate for the classroom by giving a chance to students to express their ideas (Pitzer & Skinner, 2017) and showing positive and respective attitudes toward their opinions (Kane & Staiger, 2012). In contrast to individual feedback, however, the teacher-generated supportive climate is known to be a positive predictor of student motivation rather than achievement (Lipowsky et al., 2009).

IBL demands high teacher-student interactions and constructive feedback compared to traditional science lessons since during the inquiry process, teachers are expected to facilitate IBL by addressing conceptual understanding and inquiry procedure (Kang, 2022). Accordingly, teacher support and feedback during IBL may increase students' understanding of scientific knowledge and the inquiry process (Kang, 2022). Unfortunately, during science lessons, boys often received more support and encouragement from their teachers than girls (Archer et al., 2020), but those interactions were found to be strong predictors of STEM aspirations (OECD, 2016b). Similarly, since previous studies showed that boys have participated in IBL more actively, it is plausible to assume that boys may get more feedback and support from their teachers during IBL which may lead them to pursue STEM careers more than girls. However, although previous research explored the effects of teachers on gender-science stereotypes (e.g. the presence of female teachers in STEM subjects (Riegle-Crumb et al., 2017)) and STEM aspirations (e.g. female teachers as a role model (Chen et al., 2020)), the impact of teacher-student interactions on STEM aspirations is often neglected.

IBL and Equity in Finnish Context

To explore the research questions mentioned in the "Introduction," this study used a sample representing Finnish 15-year-old students from PISA 2015 data. Similar to other European countries, the Finnish National Core Curriculum for Basic Education including lower secondary education emphasizes the implementation of IBL for science education (Finnish National Agency for Education [FNAE], 2014). For instance, concerning the learning environment and working methods of biology for 7th to 9th grades, it states that "The objective is that the pupils get an opportunity to work in diverse learning environments that support an inquiry-based approach to biology at school" (p. 401). According to Lavonen and Laaksonen (2009), Finnish secondary science teachers often implemented inquiry activities in a traditional way while students wanted more open inquiry practices (Juuti et al., 2010). Kang and Keinonen (2018) also found that Finnish students often experienced a more structured form of inquiry rather than open inquiry practices. However, Lavonen and Juuti (2016) describe that IBL in the Finnish context is not a cookbook experiment in which students merely follow some procedures given by teachers. Rather, it is more like guided inquiry in which students explore their ideas with a teacher's proper guidance and rich communication between the teacher and students during the IBL processes.

In addition to IBL, equity and equality are clearly stated not only in the general aim of secondary education but also often in the science curricula such as "Equity and equality are promoted by offering the pupils opportunities for applying physics in different contexts …" (p. 418). However, there is a significant incongruity between science performance and STEM aspirations concerning gender, as Finnish girls continuously have outperformed boys in science, but their self-efficacy and interest in science were much lower than Finnish boys at age 15 (OECD, 2016a, 2019). Interestingly, unlike other science motivations, the Finnish girls and boys presented similar levels of outcome expectations and STEM career aspirations (OECD, 2016a). However, there still are gender differences depending on the fields of STEM as girls preferred to work as health professionals three times higher than boys, while boys pursued engineering professionals four times higher than girls. That is, although girls and boys perceive learning science as useful to their future careers, they aim for different STEM paths biased by gender. This difference may stem from different preferences in science between gender during the lower secondary period as Finnish girls indicated a higher interest in biology while Finnish boys preferred physics and chemistry at age 13 (Kang et al., 2019b). A 3-year longitudinal Finnish lower secondary study (Kang et al., 2021b) also revealed gender differences in relationships between interest, self-efficacy, and career aspirations. According to the results, while interest played a major role in promoting girls' STEM career aspirations, self-efficacy had no relationship with their career goal development in STEM. Especially, girls' interest at the age of 16 indicated a much stronger association with career aspirations compared to its relationship at the age of 13. Concerning the gender disparities in STEM, they recommended that since females consider their STEM careers based on inner satisfaction rather than comparison with others, science educators should consider more student-centered approaches such as IBL fostering better environments in which girls can engage in science activities.

Also, Finland as a nation is well known for its gender equity, but at the same time, paradoxically, "Finland has one of the world's largest gender gaps in college degrees in STEM fields" (Stoet & Geary, 2018, p. 581). Given the fact that school curricula in many countries including Finland highlight the importance of (1) conducting IBL in fostering science skills and motivation and (2) promoting education equity between gender, the findings of this study are informative for the international science education community to gain a deeper understanding of the mechanisms that may contribute on students' successful learning experiences from IBL and its contribution on education equity. Specifically, this study might be useful for those educators who are trying to find ways to support both science career aspirations and educational equity in an optimal way.

Method

Sample

To explore the hypothesized model in Fig. 1, data collected from PISA 2015 was used. OECD has implemented the PISA test triennially targeting 15-year-old students from more than 70 countries since 2000 (OECD, 2016a). While the PISA assesses the students' performance in reading, mathematics, and science, it focuses on one of the three subjects for each cycle and collects a variety of information related to the subject. Accordingly, science has been the main subject in 2006 and 2015 PISA cycles and the data has been rigorously used in analyzing science education policy and practice. The sample was collected using a stratified two-stage sampling design consisting of 5782 Finnish students representing 9th graders of the population and 48.8% were female.

Measures

Concerning school science experiences, the three PISA scales—IBL, feedback, and support—were chosen. As reviewed, IBL practices at school can be biased by gender. Also, while the IBL activities demand active teacher-student interactions, these practices could be affected by student gender. Thus, we chose the three PISA 2015 scales representing students' IBL experiences and the teacher-student interactions during science lessons. Concerning the IBL experiences, six PISA IBL items were chosen that were directly related to IBL as shown in Table 4. Regarding teacher-student interactions, the PISA feedback and support scales were selected. The feedback-related construct measured personal interactions between a teacher and a student using five items such as "The teacher tells me how I am performing in this course." On the other hand, the support-related construct measured teachers' effort on maintaining a conducive learning environment, for instance, "The teacher shows an interest in every student's learning." using five items. For these school science experiences constructs, students were asked to answer among four scales (1. Never or hardly ever, 2. In some lessons, 3. In most lessons, and 4. In all lessons).

Also, three motivation scales—interest, self-efficacy, and outcome expectations—were selected which mediate the effects of learning experience on career aspirations according to the SCCT model. As described in Table 1, interest in science was measured using 5 items that present students' enjoyment of learning science. Second, outcome expectation toward science was measured using 4 items indicating students' intention to learn science subjects for extrinsic rewards such as a future career. Lastly, self-efficacy in science was measured using 8 items. Students answered with a four-point Likert scale with the response from one to four and these answers were coded or recoded so that positive scores could indicate higher levels of agreement with the asked science motivations.

Lastly, the STEM aspiration variable was created to represent students' career goals in STEM fields. For each cycle, PISA asks students "what kind of job do you expect to have when you are about 30 years old." Following the classification of science-related careers in the PISA 2015 analysis, four groups of science-related occupations (science and engineering professionals, health professionals, ICT professionals, and science technicians and associate professionals) were coded as 1 representing students' STEM aspirations and all the other careers as 0 (1 = 17%, 0 = 83%) (see OECD, 2016a, pp. 282–283).

Data Analysis

In order to test the study hypotheses, a different type of structural equation modeling (SEM) was carried out using Mplus 7.4 software (Muthén & Muthén, 2012). Since the main outcome variable, STEM aspirations, was dichotomous, the weighted least square with adjustment in mean and variance (WLSMV) estimator which is designed for analyzing categorical observed data (Li, 2016) was used and pairwise deletion was applied as a default of the WLSMV estimator. Concerning reliability and validity of the latent core constructs, first, the measurement model was tested by confirmatory factor analysis (CFA). Although each construct used in this study was theoretically well defined by other studies (e.g. Jerrim et al., 2022; OECD, 2016a), exploratory factor analysis (EFA) using principal axis factoring with varimax rotation also was implemented to examine the constructs' internal reliability (Wang & Wang, 2012). Then, convergent and discriminant validity was measured following Fornell and Larcker (1981). According to Fornell and Larcker, the average variance extracted (AVE) value of each construct should be 0.5 or higher (convergent validity); a squared root value of AVE for each latent construct should be higher than each constructs' highest correlation (discriminant validity). Also, a value of composite reliability (CR) for each construct was measured to confirm internal consistency which should be 0.7 or above.

Table 1 Description table of the three science motivation constructs

Item	Description
Interest	
INT1	I have fun when I am learning <broad science>
INT2	I like reading about <broad science> topics
INT3	I am happy working on <broad science> topics
INT4	I enjoy acquiring new knowledge in <broad science>
INT5	I am interested in learning about <broad science>
Outcome expectation	
OE1	Making an effort in my <school science> subject(s) is worth it because this will help me in the work I want to do later
OE2	What I learn in my <school science> subject(s) is important for me because I need this for what I want to do later on
OE3	Studying my <school science> subject(s) is worthwhile for me because what I learn will improve my career prospects
OE4	Many things I learn in my <school science> subject(s) will help me to get a job
Self-efficacy	
SE1	Recognize the science question that underlies a newspaper report on a health issue
SE2	Explain why earthquakes occur more frequently in some areas than in others
SE3	Describe the role of antibiotics in the treatment of disease
SE4	Identify the science question associated with the disposal of garbage
SE5	Predict how changes to an environment will affect the survival of certain species
SE6	Interpret the scientific information provided on the labeling of food items
SE7	Discuss how new evidence can lead you to change your understanding about the possibility of life on Mars
SE8	Identify the better of two explanations for the formation of acid rain

After confirming reliability and validity, multiple group structural equation modeling (MGSEM) was conducted to test statistical group differences observed in the structural parameters of the hypothesis model in Fig. 1. MGSEM is carried out in a series of steps (Byrne, 2013). First, factorial invariance is tested to check whether the hypothesized statistical model measures the same phenomenon in different groups. For this, the unconstrained model and measurement invariance model are compared and, if there are no statistical differences between the two models, factorial invariance across the groups is confirmed. Second, the measurement invariance model and fully constrained model are compared to examine group differences in any paths in the hypothesized model (metric invariance). If significant differences are found between the two models, it is evidence of the differences across the groups. Concerning these model comparisons, as recommended by Cheung and Rensvold (2002), two indices were evaluated, ΔCFI and ΔRMSEA. To confirm the invariance between models, a value of ΔCFI should be equal to or smaller than 0.01 and a value of ΔRMSEA should be equal to or less than 0.015. And, finally, using the Wald test, coefficients of each path for each group are measured and compared to examine statistical differences between the groups in any specific structural path.

Students' final weights included in the PISA dataset were incorporated to correct for selection bias (Asparouhov, 2005) and students' school ID was used to account for clustering effects using Type = Complex option in Mplus. Also, students' socio-economic status (the PISA Index of ESCS (economic, social, and cultural status)) and immigration backgrounds (first and second generation of immigrants) were included in SEM analysis. In examining the goodness of model fit for SEM, traditional cut-off values were considered: the root mean square error of approximation (RMSEA) below 0.05, standardized root mean square residual (SRMR) below 0.08, and comparative fit index (CFI) and Tucker-Lewis index (TLI) above 0.90 (Wang & Wang, 2012). Concerning the high complexity of the statistical model and the large sample size, statistical significance was considered at the level of 0.01.

Result

Reliability and Validity of the Constructs

As shown in Table 2, EFA results showed that the measurement items were loaded onto the expected factors without any significant loadings onto other factors, and CFA results indicated that all items were loaded 0.49 or higher onto the expected factors. Also, all values of Cronbach α were higher than 0.82. In addition, as presented in Table 3, the CRs of each construct were all higher than 0.7 which indicated sufficient reliability of the model. Also, the AVEs were all higher than 0.5 which ensured convergent validity. In addition, the squared root of AVE for each construct was higher than each construct's highest correlation which confirmed discriminant validity. Accordingly, these results demonstrated satisfactory reliability and validity of the latent constructs.

Gender Differences in Science Experience Concerning IBL, Feedback, and Support

Concerning IBL experiences, the result showed that female and male students indicated similar mean values concerning the time students spent doing experiments and drawing conclusions from an experiment as shown in Table 4. However, male students presented higher average values than females in argumentation and debate about investigations as well as designing their own experiments and testing their investigation ideas. Figure 2 clearly shows that while the frequencies of the practical IBL activities were similar between gender, females marked "never" much higher than boys in the planning and discussion parts of IBL. Thus, based on these results we can assume that although female and male students stayed at the same laboratory to practice inquiry as they perceived the similar frequency of these activities, males were likely to be more active in planning, designing, and arguing about the investigation while females were likely to become more passive in these IBL process as females were less aware of these activities.

Table 2 Results of EFA and CFA as well as Cronbach α

Factor	EFA	CFA (S.E.)	Mean (S.D.)	Cronbach α
Inquiry-based learning				0.82
IBL1	0.57	0.49 (0.02)	1.94 (0.78)	
IBL2	0.76	0.73 (0.01)	1.68 (0.81)	
IBL3	0.62	0.55 (0.01)	2.23 (0.85)	
IBL4	0.73	0.67 (0.02)	1.28 (0.63)	
IBL5	0.80	0.79 (0.01)	1.59 (0.78)	
IBL6	0.73	0.70 (0.01)	1.58 (0.78)	
Feedback				0.93
FB1	0.81	0.75 (0.01)	1.86 (0.76)	
FB2	0.85	0.83 (0.01)	1.81 (0.79)	
FB3	0.89	0.88 (0.01)	1.81 (0.80)	
FB4	0.90	0.91 (0.01)	1.86 (0.80)	
FB5	0.87	0.88 (0.01)	1.88 (0.82)	
Support				0.91
SP1	0.77	0.75 (0.01)	3.01 (0.87)	
SP2	0.86	0.86 (0.01)	3.29 (0.81)	
SP3	0.89	0.90 (0.01)	3.35 (0.77)	
SP4	0.82	0.79 (0.01)	3.05 (0.88)	
SP5	0.81	0.76 (0.01)	3.12 (0.88)	
Self-efficacy				0.89
SE1	0.72	0.69 (0.01)	2.71 (0.84)	
SE2	0.65	0.61 (0.01)	3.22 (0.82)	
SE3	0.69	0.65 (0.01)	2.65 (0.91)	
SE4	0.78	0.78 (0.01)	2.63 (0.85)	
SE5	0.77	0.78 (0.01)	2.67 (0.85)	
SE6	0.74	0.71 (0.01)	2.79 (0.85)	
SE7	0.76	0.74 (0.01)	2.65 (0.90)	
SE8	0.71	0.71 (0.01)	2.35 (0.93)	
Outcome expectation				0.93
OE1	0.87	0.86 (0.01)	2.78 (0.86)	
OE2	0.86	0.88 (0.01)	2.84 (0.80)	
OE3	0.89	0.91 (0.01)	2.78 (0.83)	
OE4	0.89	0.89 (0.01)	2.75 (0.82)	
Interest				0.95
INT1	0.82	0.81 (0.01)	2.67 (0.79)	
INT2	0.87	0.90 (0.00)	2.57 (0.82)	
INT3	0.87	0.90 (0.01)	2.47 (0.79)	
INT4	0.86	0.90 (0.01)	2.49 (0.80)	
INT5	0.85	0.89 (0.01)	2.63 (0.81)	

Table 3 Reliability, validity, and correlation between the core constructs

	CR	AVE	IBL	FB	SP	SE	OE	INT
IBL	0.88	0.56	**0.75**					
FB	0.95	0.81	0.34	**0.90**				
SP	0.93	0.73	0.21	0.25	**0.85**			
SE	0.91	0.57	0.18	0.12	0.13	**0.76**		
OE	0.95	0.84	0.15	0.07	0.16	0.32	**0.91**	
INT	0.96	0.84	0.13	0.13	0.29	0.43	0.43	**0.92**

Note. Model fit: CFI = 0.960, TLI = 0.956, RMSEA = 0.034 (90% CI 0.033 and 0.036). Correlations of all latent constructs are $p < 0.001$ except SE and SP ($p = 0.008$). Bold indicates squared root values of AVE.

Also, the result in Table 4 shows that male students got more personal feedback from the teachers concerning their performance, strengths, and improvement areas in science than females whereas both groups indicated a similar supportive climate from their teachers. Given that the correlation of IBL with feedback (0.34) was higher than support (0.21) in Table 3, IBL experiences were more likely to be associated with teachers' personal feedback than their general support during lessons.

Gender Differences in the Relationships Between the School Science Experiences, Science Motivations, and STEM Aspirations

Before conducting the MGSEM, factorial invariance was measured, first by comparing unconstrained and measurement invariance models. However, the fit of the hypothesized unconstrained model in Fig. 1 was not satisfactory due to several insignificant correlations between the factors for both girls and boys. Thus, based on the modification indices offered by Mplus, the final model was established as presented in Fig. 3. With this model, the invariance test was conducted and the results showed that although a chi-square difference test indicated a significant difference ($p < 0.05$), CFI and RMSEA values did not change significantly (ΔCFI and ΔRMSEA < 0.01) between unconstrained and measurement invariance models as shown in Table 5. Thus, this result indicated that the factor loadings were invariant across gender. Then, the metric invariance was tested by comparing the measurement invariance model and the fully constrained model. According to the result, both the chi-square difference test and model fits indicated significant differences between the two models (ΔCFI = 0.77); this result presented that some path coefficients might be variant across gender.

Accordingly, the paths presented in Fig. 3 were constrained one by one to investigate on which paths female and male indicated differences. As shown in Fig. 3 and Table 6, the results of Wald tests confirmed the four paths that had gender differences. Especially, females' IBL indicated stronger positive relationships with self-efficacy than males; females' self-efficacy indicated a higher correlation with interest than males; the relationship between interest and STEM aspirations was statistically significant for girls whereas it was insignificant for boys. Therefore, we can assume that the effect of IBL on STEM is mediated by self-efficacy and interest in the case of girls.

Table 4 Gender differences in school science experiences

Item	Description	Female	Male	*U*
Inquiry-based learning				
IBL1	Students spend time in the laboratory doing practical experiments	1.91 (0.74)	1.97 (0.81)	– 2.21
IBL2	Students are required to argue about science questions	1.53 (0.73)	1.84 (0.86)	-14.53^{***}
IBL3	Students are asked to draw conclusions from an experiment they have conducted	2.22 (0.83)	2.24 (0.87)	– 0.82
IBL4	Students are allowed to design their own experiments	1.17 (0.48)	1.38 (0.73)	-11.92^{***}
IBL5	There is a class debate about investigations	1.44 (0.67)	1.73 (0.85)	-13.30^{***}
IBL6	Students are asked to do an investigation to test ideas	1.43 (0.68)	1.72 (0.84)	-13.65^{***}
Feedback				
FB1	The teacher tells me how I am performing in this course	1.73 (0.74)	1.98 (0.75)	-12.92^{***}
FB2	The teacher gives me feedback on my strengths < school science > subject	1.66 (0.76)	1.96 (0.79)	-14.65^{***}
FB3	The teacher tells me in which areas I can still improve	1.65 (0.76)	1.97 (0.81)	-15.55^{***}
FB4	The teacher tells me how I can improve my performance	1.71 (0.77)	2 (0.81)	-13.72^{***}
FB5	The teacher advises me on how to reach my learning goals	1.75 (0.79)	2.01 (0.83)	-12.21^{***}
Support				
SP1	The teacher shows an interest in every student's learning	2.97 (0.88)	3.05 (0.85)	-2.88^{**}
SP2	The teacher gives extra help when students need it	3.26 (0.82)	3.31 (0.79)	– 2.40
SP3	The teacher helps students with their learning	3.33 (0.79)	3.37 (0.76)	– 1.70
SP4	The teacher continues teaching until the students understand	3.04 (0.9)	3.07 (0.86)	– 1.07
SP5	Teacher gives an opportunity to express opinions	3.09 (0.9)	3.14 (0.85)	– 2.05

Note. $^{***} < 0.001$, $^{**} < 0.005$.

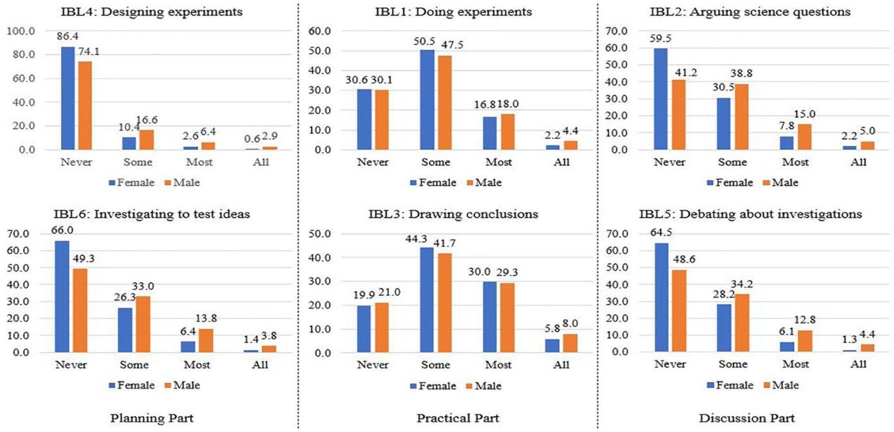

Fig. 2 Frequency values of the four Likert scales for the six IBL-related activities

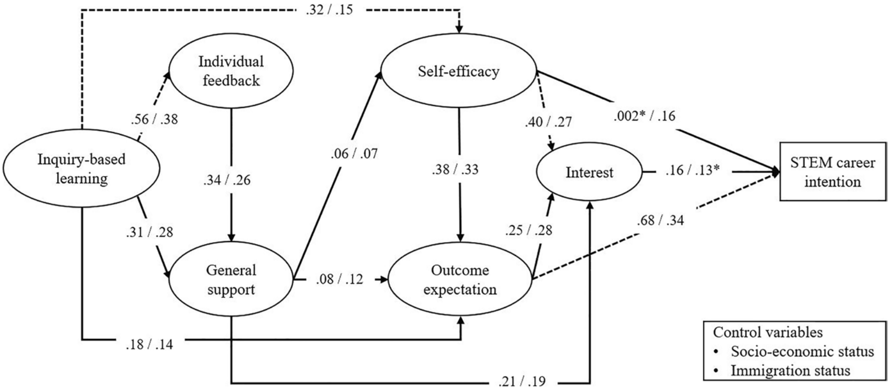

Fig. 3 Final model investigating the relationships between the core factors. Note. Unstandardized coefficients are presented. Dashed lines indicated significant gender differences in the specific paths. All path coefficients are significant at the level of 0.01 except for two * marked paths

Table 5 Result of the model invariance test

Model	χ^2 (*df*)	CFI	RMSEA (90% CI)	χ^2 (*df*) difference	ΔCFI	ΔRMSEA
1. Unconstrained	3450.29 (1196)	0.926	0.026 (0.025, 0.027)			
2. Measurement invariance	3393.22 (1223)	0.928	0.025 (0.024, 0.026)	67.71 (27), $p<0.05$	0.002	0.001
3. Fully constrained	5850.34 (1347)	0.851	0.034 (0.033, 0.035)	1625.46 (124), $p<0.05$	0.077	0.009

Also, for both genders, while outcome expectation indicated a higher correlation with STEM aspirations than interest or self-efficacy, the coefficient for girls was significantly higher than for boys. Interestingly, the effect of self-efficacy on STEM aspirations was statistically significant for boys while the coefficient for the girls was statistically non-significant. On the other hand, the relationship between interest and STEM aspirations was statistically significant for girls whereas it was insignificant for boys. Thus, the result indicated that in pursuing STEM careers, self-efficacy may be more important for boys, while interest is more pivotal for girls.

Considering control variables, students' socio-economic status indicated significant relationships with all factors except feedback whereas immigrant background indicated only a significant relationship with feedback ($b = 0.07$ ($p < 0.001$) and $b = 0.06$ ($p < 0.001$) for the first- and second-generation immigrants, respectively). Thus, the result indicated that feedback was given more to the students with immigration backgrounds while socio-economic status had no associations with feedback.

Finally, indirect effects of IBL, feedback, and support on STEM career aspirations were measured. As presented in Table 7, IBL and support affected STEM career aspirations positively for both females and males, but effect sizes were small; the effect of feedback on STEM aspirations was negligible; the coefficients of IBL and support were slightly higher for girls than boys.

Discussion

At school, the use of collaborative learning and small group activities are encouraged due to their potential to foster successful problem-solving. Especially, collaboration in a small group has always played an important role in science education

Table 6 Wald (χ^2) test results and path coefficients for each group

Constrained path	χ^2 (*df*)	Girls (S.E.)	Boys (S.E.)
1. Interest → STEM career	0.12 (1)	0.10^{**} (0.04)	0.09 (0.04)
2. Outcome → STEM career	$45.42 (1)^{***}$	0.48^{***} (0.02)	0.25^{***} (0.03)
3. Efficacy → STEM career	3.35 (1)	0.001 (0.03)	0.09^{*} (0.03)
4. Outcome → interest	2.98 (1)	0.29^{***} (0.02)	0.31^{***} (0.02)
5. Efficacy → interest	$17.43 (1)^{***}$	0.35^{***} (0.02)	0.24^{***} (0.02)
6. Support → interest	1.05 (1)	0.25^{***} (0.02)	0.20^{***} (0.02)
7. Efficacy → outcome	1.59 (1)	0.29^{***} (0.02)	0.26^{***} (0.02)
8. Support → outcome	1.61 (1)	0.08^{**} (0.02)	0.12^{***} (0.02)
9. IBL → outcome	0.59 (1)	0.09^{***} (0.02)	0.10^{***} (0.02)
10. Support → efficacy	0.32 (1)	0.08^{**} (0.03)	0.09^{***} (0.02)
11. IBL → efficacy	$11.74 (1)^{***}$	0.21^{***} (0.03)	0.13^{***} (0.02)
12. Feedback → support	3.00 (1)	0.26^{***} (0.03)	0.21^{***} (0.02)
13. IBL → support	0.31 (1)	0.15^{***} (0.03)	0.19^{***} (0.02)
14. IBL → feedback	$14.61 (1)^{**}$	0.36^{***} (0.02)	0.33^{***} (0.02)

Note. $^{***} < 0.001$, $^{**} < 0.005$. Standardized path coefficients.

Table 7 Indirect effects of IBL, feedback, and support on STEM career aspirations

	Girls		Boys	
	β	SE	β	SE
STEM career by IBL	0.10^{***}	0.01	0.07^{***}	0.01
STEM career by feedback	0.02^{***}	<0.01	0.01^{***}	<0.01
STEM career by support	0.08^{***}	0.02	0.07^{***}	0.01

Note. *** < 0.001. Standardized path coefficients.

because in nature it involves a group laboratory work as a part of inquiry to solve scientific problems. Accordingly, IBL has been placed in the center of science curricula in many countries including Finland and found to be a significant predictor of students' interest in science studies and work. However, IBL activities have been also criticized as a possible contributor to gender disparities in STEM because boys often get more opportunities to engage in hands-on problem solving while girls participate more in passive activities; boys were more likely to get support and feedback from their teachers (Tindall & Hamil, 2004; Wieselmann et al., 2020). However, the previous findings were based on small-scale observational studies, and few investigated this phenomenon quantitatively. Also, while much research showed a linear relationship between IBL experiences and STEM aspirations, only a few used a rigorous theoretical framework such as SCCT in investigating the relationship. Thus, this study aimed to explore the girls' and boys' perceptions of students' participation in IBL activities and teacher support during science lessons using ILSA data. In addition, this study investigated the effects of IBL on students' aspirations and gender effects in this relationship using the SCCT framework.

Gendered Patterns in IBL Experiences at School and Their Effect

According to the result, although 15-year-old girls and boys indicated a similar perception of students' participation in the laboratory work in Finland, boys were more positive than girls in perception of planning and designing of inquiry and auguring about science questions than girls. This result is in line with the study by Lee and Burkam (1996) showing that the eighth graders both girls and boys experienced similar frequencies of IBL at school in the USA, but our study further showed a more detailed picture concerning the gender disparities in IBL engagement by taking a different type of IBL activities into account. One of the reasons that caused the gender differences in perception of students' inquiry learning experiences is probably due to their different roles played during the inquiry processes. According to previous studies, the school science practices tend to be gendered in a way that boys often take control of the process and do not collaborate with girls (e.g. Wieselmann et al., 2020). Accordingly, girls are prone to behave passively and may remember the certain type of science practices less than boys in this kind of retrospective study. However, these gendered school science learning experiences may result in a negative relationship between the IBL practice and science achievement. For instance, previous Finnish

PISA studies showed that doing laboratory work and drawing conclusions from the experiments (that both genders perceived equally in this study) are typical scientific practices at Finnish secondary schools (Kang & Keinonen, 2018). On the other hand, the other four types of IBL activities (that were more perceived by boys in this study) indicated a negative relationship with achievement, but a positive relationship with interest in many countries including Finland (Aditomo & Klieme, 2020; Kang & Keinonen, 2018). Thus, this study together with the previous PISA studies shows that the inquiry learning experiences that do not offer equal chances for participation between gender will not guarantee positive learning outcomes.

Interestingly, while girls perceived fewer IBL learning experiences, especially those that gave higher autonomy to students, the positive association of IBL with self-efficacy was much greater for girls, and the associations with outcome expectations and STEM aspirations were also slightly higher for girls than boys. That is, the girls' learning experiences were more efficient for enhancing their science motivations than their counterpart boys according to the hypothesized SCCT model. This result is in line with the study by Baker (2013) underlining a genuine inquiry experience as a key factor fostering girls' interest in science. Concerning the Finnish females' low interest and self-efficacy in science (OECD, 2016a) as well as low engagement in STEM careers (Stoet & Geary, 2018), the open inquiry culture should be developed at the secondary school in Finland so that girls' science learning experiences at school could be a source of their STEM career aspirations in future. Fortunately, open-ended IBL activities are gradually adopted as a part of school core curricula, for instance, in Finland, the Netherlands, Korea, and Israel (Kang, 2022). However, as our results show, when conducting high autonomous IBL activities or open inquiry, equal participation between gender should be taken into account, especially for girls who are likely to be marginalized during the activities. Although in this study the effect size of IBL on STEM aspirations was small, we anticipate that when the girls' perception of their participation in the high autonomous IBL activities increases, the relationship between IBL and science motivations including STEM aspirations might be enhanced.

Gender Differences in Having Feedback During Science Lessons and Their Implication

In addition to IBL, girls indicated less individual feedback from their science teachers than boys although both girls and boys perceived a similar supportive climate during science lessons. Also, the result shows that while a supportive climate indicated positive relationships with motivation factors, feedback was not associated with any science motivations. As is known, Finnish education emphasizes the importance of seeing individuals' needs and responding to learner diversity (FNAE, 2014). Also, there is systematic support at the national, municipal, and school levels to find and help students who are in need (Harju & Niemi, 2018). At the same time, in pre and in-service teacher education, making a classroom environment conducive to learning for all students is emphasized (Niemi et al., 2016). As shown in this study, however, although the level of support that students perceived as a group was very high, individual feedback that students had from their science teachers indicated gender inequity. One possible reason that boys got more feedback was that, in the Finnish context, feedback was likely to be given more to underachievers rather than to high achievers. That is, since Finnish boys were less knowledgeable than girls in science (OECD, 2016a), they were likely to get more feedback from the teacher to improve their science performance (Wisniewski et al., 2020). PISA 2015 also reported that "More perceived feedback is also associated with poorer performance in science, probably because low-performing students need and receive more feedback than better-performing students" (OECD, 2016b, p. 66) for many participating countries. Also, this study shows that students with immigration backgrounds got more feedback than the Finnish native students probably due to their lacking knowledge of language and science. Interestingly, the level of feedback was not related to students' socio-economic status. That is, in Finland, since there are negligible differences in science achievement among different socio-economic backgrounds (OECD, 2016a), feedback is equally distributed regardless of the student's status. Since this study as well as other PISA studies is correlational in nature, it is not possible to claim the causal relationship between feedback and science achievement. However, as the previous study indicates that feedback is more used or useful for better academic performance (Hattie & Timperley, 2007; Wisniewski et al., 2020), we can generally assume that feedback was given to the low achievers in the Finnish context and that may be the reason why boys got more feedback compared to girls during science lessons. At the same time, it might be the reason why the feedback experiences indicated a very small or insignificant association with science motivations. Although feedback did not present any connections with science motivations in this study, however, positive and specific feedback on things girls can control such as effort or behaviors can improve their self-efficacy and performance (e.g. Kim et al., 2007); thus, the result should be carefully discussed.

Another interesting finding from this study is that while feedback was moderately and positively associated with IBL, the relationship between IBL and feedback was significantly higher for girls than boys. In other words, a chance to get feedback is higher for girls than boys when they participate in IBL. Concerning that feedback is negatively associated with performance in the Finnish context, this result implies that the teacher may undervalue girls' abilities in science so that girls are likely to get more feedback than boys when they participate in IBL at school. The previous studies showed that while girls outperformed boys in science, their abilities were often undervalued by teachers (Tindall & Hamil, 2004; Wang & Degol, 2017) and peers (Bloodhart et al., 2020) and it may also happen during IBL processes at school according to the result of this study. Kang (2022) argued that when students experienced open inquiry (high autonomous IBL activities), teachers often called a few selected students who were more active or good at science, students could not get good marks after open inquiry, and they easily got nervous during the inquiry process which in the end eroded teacher-student relationships. Therefore, if a teacher undervalues girls' science performance and girls could not be an active participants, it is plausible to assume that girls may less get attention and be called by the teacher while boys may get more attention and be encouraged to keep up their work which, eventually, widen gender disparities in IBL engagement.

Unfortunately, even in the classrooms of experienced teachers concerned with gender equity, girls' engagement in IBL activities such as manipulating laboratory equipment was much less than boys (e.g. Jovanovic & Steinbach King, 1998) and much research has indicated that teachers have limited knowledge and experience with gender inclusivity (Brotman & Moore, 2008). Therefore, gender-inclusive science curricula dealing with for instance gendered patterns in communication between the teacher and students, and engagements in science practices as well as teachers' views of girls' science performance should be thoroughly addressed during the pre and in-service teacher professional development to have them a better understanding of gender-responsive science teaching (Stapleton, 2015) and to increase girls' learning experiences in high autonomous inquiry activities that will foster their self-efficacy, outcome expectations, and STEM career aspirations.

Limitation and Suggestions

By using the secondary cross-sectional data, this study includes some limitations. First, as reviewed, although previous research indicated that a critical period of young students' STEM aspirations may begin in the upper primary school years, this study only used secondary school student data. Thus, for future research, it would be beneficial to investigate whether the gendered IBL practices could be also found in primary education at large and how it affects students' STEM aspirations.

Second, the feedback and support constructs used in this study were not directly linked to the IBL situation but focused more on overall science lessons. Therefore, it is recommended to use a measurement directly linked to the IBL context in investigating the relationships between IBL and teacher-student interactions and compare the result with this study.

In addition, the items related to IBL and support focused on collective experiences (e.g. STUDENTS spend time in the laboratory …) while other items measured personal experiences (e.g. the teacher tells ME how I am performing …). Although the collective perspective still offers valuable information, it might be more accurate if the IBL and support items used the first-person point of view in measuring individual perception. Thus, we recommend revising the questions in a way that all the questions use the personal viewpoint for future studies examining group differences such as gender or race.

Also, as already mentioned, this is a retrospective correlation study, so it is not possible to make any conclusion in terms of causal relationships between the constructs. Therefore, for future studies, interviews or classroom observation as well as follow-up longitudinal studies are recommended.

Lastly, although we included students' immigration status to control the effect of ethnic background which also is known to be an IBL engagement predictor, since Finland is a relatively ethnically homogeneous country, it may be hard to say that the effect of ethnic background is rigorously considered. Thus, similar studies are recommended for the countries where there are many diverse ethnic groups among the student population in school.

Conclusion

While IBL has become essential for science education and has been adopted in other school subjects as a form of collaborative learning, less attention has been given to issues of equity in the use of IBL. The findings of this study, therefore, shed light on how IBL has been practiced at school in terms of gender perspectives and how gendered patterns can affect science motivations. Also, this study contributes to the previous literature on the gender effect on the relationship between IBL and science performance by revealing the gendered practices of IBL and its relationships with other science experiences and motivations.

Funding Open access funding provided by University of Eastern Finland (UEF) including Kuopio University Hospital.

Declarations

Ethical Approval and Consent to Participate This paper used data that are openly available for educational and research purposes collected by the Organisation for Economic Co-operation and Development (OECD) (seehttp://www.oecd.org/pisa/data/2015database). Thus, ethical approval is not applicable for this article.

References

Aditomo, A., & Klieme, E. (2020). Forms of inquiry-based science instruction and their relations with learning outcomes: Evidence from high and low-performing education systems. *International Journal of Science Education, 42*(4), 504–525.

Almukhambetova, A., Torrano, D. H., & Nam, A. (2021). Fixing the leaky pipeline for talented women in STEM. *International Journal of Science and Mathematics Education*. https://doi.org/10.1007/s10763-021-10239-1

Archer, L., Moote, J., MacLeod, E., Francis, B., & DeWitt, J. (2020). *ASPIRES 2: Young people's science and career aspirations, age 10–19*. UCL Institute of Education.

Asparouhov, T. (2005). Sampling weights in latent variable modeling. *Structural Equation Modeling, 12*(3), 411–434.

Baker, D. (2013). What works: Using curriculum and pedagogy to increase girls' interest and participation in science. *Theory into Practice, 52*(1), 14–20.

Bloodhart, B., Balgopal, M. M., Casper, A. M. A., Sample McMeeking, L. B., & Fischer, E. V. (2020). Outperforming yet undervalued: Undergraduate women in STEM. *PLoS One, 15*(6), e0234685.

Brotman, J. S., & Moore, F. M. (2008). Girls and science: A review of four themes in the science education literature. *Journal of Research in Science Teaching, 45*(9), 971–1002.

Byrne, B. M. (2013). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). Taylor and Francis. https://doi.org/10.4324/9780203805534

Cairns, D., & Areepattamannil, S. (2019). Exploring the relations of inquiry-based teaching to science achievement and dispositions in 54 countries. *Research in Science Education, 49*(1), 1–23.
Çalık, M. (2013). Effect of technology-embedded scientific inquiry on senior science student teachers' self-efficacy. *Eurasia Journal of Mathematics, Science and Technology Education, 9*(3), 223–232.
Chen, C., Sonnert, G., & Sadler, P. M. (2020). The effect of first high school science teacher's gender and gender matching on students' science identity in college. *Science Education, 104*(1), 75–99.
Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233–255.
Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125*(6), 627–668.
Finnish National Agency for Education [FNAE]. (2014). *National core curriculum for basic education 2014*. Author.
Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39–50.
Giffi, C. A., Rodriguez, M. D., Gangula, B., Michalik, J., de la Rubia, T. D., Carbeck, J., & Cotteleer, M. J. (2015). *Advanced technologies initiative: Manufacturing & innovation* (pp. 2–70). Deloitte and Council on Competitiveness.
Harju, V., & Niemi, H. (2018). Teachers' changing work and support needs from the perspectives of school leaders and newly qualified teachers in the Finnish context. *European Journal of Teacher Education, 41*(5), 670–687.
Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.
Holmes, K., Gore, J., Smith, M., & Lloyd, A. (2018). An integrated analysis of school students' aspirations for STEM careers: Which student and school factors are most predictive? *International Journal of Science and Mathematics Education, 16*(4), 655–675.
Jeffries, D., Curtis, D. D., & Conner, L. N. (2020). Student factors influencing STEM subject choice in year 12: A structural equation model using PISA/LSAY data. *International Journal of Science and Mathematics Education, 18*(3), 441–461.
Jerrim, J., Oliver, M., & Sims, S. (2022). The relationship between inquiry-based teaching and students' achievement. New evidence from a longitudinal PISA study in England. *Learning and Instruction, 80*, 101310.
Jovanovic, J., & Steinbach King, S. (1998). Boys and girls in the performance-based science classroom: Who's doing the performing? *American Educational Research Journal, 35*, 477–496.
Juuti, K., Lavonen, J., Uitto, A., Byman, R., & Meisalo, V. (2010). Science teaching methods preferred by grade 9 students in Finland. *International Journal of Science and Mathematics Education, 8*(4), 611–632.
Kane, T., & Staiger, D. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains* [Research paper]. MET Project. Bill & Melinda Gates Foundation.
Kang, J. (2022). Interrelationship between inquiry-based learning and instructional quality in predicting science literacy. *Research in Science Education, 52*(1), 339–355.
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.
Kang, J., & Keinonen, T. (2018). The effect of student-centered approaches on students' interest and achievement in science: Relevant topic-based, open and guided inquiry-based, and discussion-based approaches. *Research in Science Education, 48*(4), 865–885.
Kang, J., Keinonen, T., Simon, S., Rannikmäe, M., Soobard, R., & Direito, I. (2019a). Scenario evaluation with relevance and interest (SERI): Development and validation of a scenario measurement tool for context-based learning. *International Journal of Science and Mathematics Education, 17*(7), 1317–1338.
Kang, J., Hense, J., Scheersoi, A., & Keinonen, T. (2019b). Gender study on the relationships between science interest and future career perspectives. *International Journal of Science Education, 41*(1), 80–101.
Kang, J., Salonen, A., Tolppanen, S., Scheersoi, A., Hense, J., Rannikmäe, M., Soobard, R., & Keinonen, T. (2021a). Effect of embedded careers education in science lessons on students' interest, awareness,

and aspirations. *International Journal of Science and Mathematics Education*. https://doi.org/10.1007/s10763-021-10238-2
Kang, J., Keinonen, T., & Salonen, A. (2021b). Role of interest and self-concept in predicting science aspirations: Gender study. *Research in Science Education, 51*(1), 513–535.
Kim, Y., Wei, Q., Xu, B., Ko, Y., & Ilieva, V. (2007). MathGirls: Toward developing girls' positive attitude and self-efficacy through pedagogical agents. In R. Luckin, K. R. Koedinger & J. Greer (Eds.), *Artificial intelligence in education: Building technology rich learning contexts that work* (Vol. 158, pp. 119–126). IOS Press.
Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*, 75–86.
Lavonen, J., & Juuti, K. (2016). Science at Finnish compulsory school. In H. Niemi, A. Toom & A. Kallioniemi (Eds.), *The miracle of education: The principles and practices of teaching and learning in Finnish schools* (pp. 125–144). Sense Publishers.
Lavonen, J., & Laaksonen, S. (2009). Context of teaching and learning school science in Finland: Reflections on PISA 2006 results. *Journal of Research in Science Teaching, 46*(8), 922–944.
Lee, V. E., & Burkam, D. T. (1996). Gender differences in middle grade science achievement: Subject domain, ability level, and course emphasis. *Science Education, 80*(6), 613–650.
Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior, 45*(1), 79–122.
Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48*(3), 936–949.
Lipko-Speed, A., Dunlosky, J., & Rawson, K. A. (2014). Does testing with feedback help grade-school children learn key concepts in science? *Journal of Applied Research in Memory and Cognition, 3*(3), 171–176.
Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E., & Reusser, K. (2009). Quality of geometry instruction and its short-term impact on students' understanding of the Pythagorean theorem. *Learning and Instruction, 19*(6), 527–537.
Ma, X., & Willms, J. D. (2004). School disciplinary climate: Characteristics and effects on eighth grade achievement. *Alberta Journal of Educational Research, 50*(2), 169–188.
Miller, D. I., Nolla, K. M., Eagly, A. H., & Uttal, D. H. (2018). The development of children's gender-science stereotypes: A meta-analysis of 5 decades of US Draw-a-Scientist studies. *Child Development, 89*(6), 1943–1955.
Moorhouse, E. A. (2017). Sex segregation by field of study and the influence of labor markets: Evidence from 39 countries. *International Journal of Comparative Sociology, 58*, 3–32.
Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (6th ed.). Authors.
National Science Board. (2015). *Revisiting the STEM workforce: A companion to science and engineering indicators 2014*. Author.
Niemi, H., Toom, A., & Kallioniemi, A. (2016). *Miracle of education: The principles and practices of teaching and learning in Finnish schools.* Sense Publishers.
Oliver, M., McConney, A., & Woods-McConney, A. (2021). The efficacy of inquiry-based instruction in science: A comparative analysis of six countries using PISA 2015. *Research in Science Education, 51*, 595–616.
Organization for Economic Cooperation and Development [OECD]. (2016a). *PISA 2015 results (volume I): excellence and equity in education*. Author.
Organization for Economic Cooperation and Development [OECD]. (2016b). *PISA 2015 results (volume II): policies and practices for successful schools*. Author.
Organization for Economic Cooperation and Development [OECD]. (2019). *PISA 2018 results (volume I): What students know and can do*. Author.
Pitzer, J., & Skinner, E. (2017). Predictors of changes in students' motivational resilience over the school year: The roles of teacher support, self-appraisals, and emotional reactivity. *International Journal of Behavioral Development, 41*(1), 15–29.
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.
Riegle-Crumb, C., Moore, C., & Buontempo, J. (2017). Shifting STEM stereotypes? Considering the role of peer and teacher gender. *Journal of Research on Adolescence, 27*(3), 492–505.

Şimşek, P., & Kabapınar, F. (2010). The effects of inquiry-based learning on elementary students' conceptual understanding of matter, scientific process skills and science attitudes. *Procedia-Social and Behavioral Sciences, 2*(2), 1190–1194.
Sithole, A., Chiyaka, E. T., McCarthy, P., Mupinga, D. M., Bucklein, B. K., & Kibirige, J. (2017). Student attraction, persistence and retention in STEM programs: Successes and continuing challenges. *Higher Education Studies, 7*(1), 46–59.
Sjøberg, S. (2017). PISA as a challenge for science education: Inherent problems and problematic results from a global assessment regime. *Revista Brasileira De Pesquisa Em Educação Em Ciências., 1*(17), 327–263.
Stapleton, S. R. (2015). Supporting teachers for race-, class-, and gender-responsive science teaching. *Cultural Studies of Science Education, 10*(2), 411–418.
Stoet, G., & Geary, D. C. (2018). The gender-equality paradox in science, technology, engineering, and mathematics education. *Psychological Science, 29*(4), 581–593.
Taskinen, P. H., Schütte, K., & Prenzel, M. (2013). Adolescents' motivation to select an academic science-related career: The role of school factors, individual interest, and science self-concept. *Educational Research and Evaluation, 19*(8), 717–733.
Tindall, T., & Hamil, B. (2004). Gender disparity in science education: The causes, consequences, and solutions. *Education, 125*(2), 282–295.
Wang, M. T., & Degol, J. L. (2017). Gender gap in science, technology, engineering, and mathematics (STEM): Current knowledge, implications for practice, policy, and future directions. *Educational Psychology Review, 29*(1), 119–140.
Wang, J., & Wang, X. (2012). *Structural equation modeling: Applications using Mplus*. Wiley.
Wieselmann, J. R., Dare, E. A., Ring-Whalen, E. A., & Roehrig, G. H. (2020). "I just do what the boys tell me": Exploring small group student interactions in an integrated STEM unit. *Journal of Research in Science Teaching, 57*(1), 112–144.
Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology, 10*, 3087.