



JOURNAL
OF BALTIC
SCIENCE
EDUCATION

ISSN 1648-3898 /Print/

ISSN 2538-7138 /Online/

Abstract. *This research explores whether classroom life (CL), textbooks (TE), and learning initiative (LI) are mediators between instruction (P) and science performance, as well as whether these mechanisms are moderated by gender. 484 eighth grade students completed the questionnaire with four subscales of P, LI, TE, and CL. For the needs of triangulation and complementarity, three focus group interviews were conducted later. Based on mediation analysis and multi-group structural equation modeling, it was found that 1) the direct effects of P on LI, P on CL, P on TE, and LI on science performance are significant, while the other direct effects are insignificant; 2) comparing to the male group, the direct effect of P on LI in the female group is larger; 3) characteristics hindering students' science learning include: the pace of a lesson is too fast, pictures and experiments are less in the textbook, and top students and low proficiency students are uncooperative. Findings expose that instruction significantly influence students' science performance, and this impact is completely mediated by students' learning initiatives. The relation between instruction and learning initiative is stronger in the female group. Textbooks can be useless in the context that instruction does not match students' learning ability.*

Keywords: *eighth-grade students, gender difference, mediation analysis, science performance*

Tao Jiang, Ji-gen Chen, Yin-yin Wu
Taizhou University, China

IMPACT OF INSTRUCTION ON SCIENCE PERFORMANCE: LEARNING INITIATIVE AS A MEDIATOR AND GENDER AS A LIMITED MODERATOR

**Tao Jiang, Ji-gen Chen,
Yin-yin Wu**

Introduction

Asian science is experiencing rapid development. The same is true for its science education. The 15 year old students in Singapore, Macao(China), Hong Kong(China), Beijing-Shanghai-Jiangsu-Guangdong(China), Beijing-Shanghai-Jiangsu-Zhejiang(China), and Korea performed well in mathematics and science tests (OECD, 2016; OECD, 2019a). Nevertheless, there are several deficiencies in Asian science education. The exam-centric education system in most of the Asian countries usually concentrates on content knowledge and neglects of nurturing students' innovative thinking, which would hamper the rise of Asian science (Lim, 2010). For example, comparing to American college students, Chinese students gained much better scores on tests of physics content knowledge but failed to maintain that advantage on tests of scientific reasoning (Bao et al., 2009). Meanwhile, it seems that some traditional advanced countries, such as the United States and the UK, also face challenges in maintaining their leading advantages in science. Despite many years of standards-based reform, the US only had minimal improvements in its science education ("The Science of Education Reform," 2006). The US students' mathematics and science achievements in the 2009 Program for International Student Assessment (PISA) test were significantly behind the Asian participating countries and other developed nations ("Change the Equation," 2011), and this status was not meliorated in the 2015 and 2018 PISA tests, especially in the mathematics test (OECD, 2016; OECD, 2019a). Although the UK students' science and mathematics performance in 2015 and 2018 PISA tests were better than the US students, the UK also needed to deal with its students losing their interests in mathematics and science during the secondary school period ("Science Education Reforms in the UK," 2012). On account of there are some important factors in school environments influencing students' science learning, this paper intends to depict a clear picture of how these factors are integrated to affect a student's science performance.

Factors Influencing Students' Science Performance in School Context

Heredity, social-economic status, the quality of schooling, investment on per student by governments, and cultural and social settings explain students' science performance. This research has no intention to depict such a big picture and only focus on exploring the mechanism of science performance production in the school context.

In school science, the stereotype is deeply entrenched that men in nature are more competitive than women (Fine & Elgar, 2017). Poor media portrayals such as women are responsible for housekeeping, and a lack of role models who are successful in a career may be responsible for that (Saujani, 2017). Therefore, the gender difference in science learning is something related to women's role definition, which human society imposes on them. Not dramatically, but gradually this stereotype has been changing. The increasing urbanization and well-educated parents may play a big role in terms of this change (Burušić et al., 2019; Gupta, 2017). By early intervention, Saujani (2017) succeeded in teaching middle school female students to write computer programs. Beaman et al. (2012) found that the policy experiment of "Female Leadership" in India could promote girls' educational attainment. Although it was found that boys did better in mathematics in the 2003 PISA test (Machin & Pekkarinen, 2008), Guiso et al. (2008) argued that this conclusion did not hold in countries with a more gender-equal culture. Coll et al. (2010) also claimed that there was little gender difference in New Zealand students' science performance in PISA 2006, yet girls' science performance was 17 scores higher than boys' performance in Thailand. PISA 2018 results again reported that 'girls outperformed boys in science by two score points on average across OECD countries' (OECD, 2019b). Nevertheless, by analyzing the "Trends in International Mathematics and Science Study" (TIMSS) 2015 data, Aşkın and Öz (2020) argued though girls outperformed boys in science in 5 Asian countries, the opposite side was true for Georgia, Italy, Lithuania, and the United States. In a global view, girls have progressed in their science performance, both boys and girls can be successful in learning science. Therefore, it is valuable to explore the mechanism that accounts for boys' and girls' success in science classrooms.

In school science, instruction and learning, and their interactions influence students' learning. Teachers who accommodate their instruction to students' learning levels could improve students' test-scores (Kremer et al., 2013). Also, strategies that prompt learner's engagement in science tasks are repeatedly emphasized by scholars. A large number of these strategies, such as enhancing teacher-student interactions (Allen et al., 2011), taking notes through the mind-mapping method (Akinoglu & Yasar, 2007), active learning (Freeman et al., 2014), direct instruction with hands-on and minds-on attributes (Cobern et al., 2010), inquiry-based reform (Sotáková et al., 2020), integrating doing, reading, writing, and talking (Webb, 2010), making connections between student's life and subject matter knowledge (Hulleman & Harackiewicz, 2009), and using learning techniques (Dunlosky et al., 2013) were reported to positively relate to students' learning gains. Conversely, in classrooms where teachers did most of the talking, student's performance in tests was remarkably low (Setati et al., 2002). However, teachers talking lot was a typical scene in mainland China's science classrooms, but it did not prevent Chinese students from getting good scores in PISA science tests. Then, it raises two important questions. First, by which means does direct instruction affect students' science learning? Second, which behaviors of direct instruction have hindered their science learning from students' viewpoints? However, previous research paid little attention to the answers of these two questions.

In the school context, textbooks are essential resources (Oates, 2014; Wilkens, 2011). It supports teaching and learning to make sure the pedagogy is structured (Reichenberg, 2016). For this purpose, it always includes the necessary tools for learning (Hanbay, 2015), such as pictures, tables, and laboratory instructions. The utilization levels of textbooks vary in countries due to their educational systems. Chinese teachers usually cover 100% subject matter knowledge (SMK) in textbooks in their lessons, because direct instruction makes it feasible by controlling time spent on a topic. While in the United States, less than 50% of high school science teachers, and less than 70% of high school mathematics teachers covered more than 75% SMK in textbooks in their lessons (Banilower et al., 2013). In England, only 10% of teachers viewed textbooks as a basis for instruction, in contrast to 70% in Singapore, and 95% in Finland (Oates, 2014). So, it seemed that a country's textbooks utilization had some kind of relationship with her PISA test score. But this relationship also could be a coincidence, it might be other factors in the school context rather than textbooks, contributing to students' test performance. Therefore, it is in urgent need of investigating whether and how science textbooks influence students' performance. If a textbook does play as an irreplaceable role in students' learning outcomes, teachers should value and maximize its utilization. If a textbook does not play a big role in students' learning, in consideration of the fact that it cost a lot of money every year, there is a voice of replacing it with the low-cost electronic textbook (Robinson, 2011). However, it is hard



to choose between the two arguments as past research has said little about the mechanism of science textbook played in students' performance with a quantitative method.

Some researchers made efforts to understand students' classroom life (Brophy, 2006; Johnson et al., 2009). Classrooms are shared social spaces where participants' personal and institutional lives are weaved together (Gieve & Miller, 2006). In institutional lives, teachers and students make efforts to resolve problems; in personal lives, they speak to each other. The personal dimension highlights the emotional connections of participants, while the emotional characteristics of classrooms influence students' learning outcomes (Fraser, 1987). For example, mutual respect between teachers and students prompted students' engaging in tasks (Matsumura et al., 2008). For teachers, more interactions fewer interventions improved their students' achievements (Brophy, 2006; Djigic & Stojiljkovic, 2011). For students' community, more respect less public competition enhanced their learning outcomes (Pierce, 1994). Recently, it was reported that school bullying has shaken the public's trust in schools as a place of social learning and development (Sachs & Mellor, 2005). Teachers also experienced "culture shock" and burnout in schools where their students faced too much violence on the street (Rushton, 2000). If the minimum level of students' safety is unrealizable, there will be no authentic students' engagement in tasks. Fortunately, conditions like that are rare, students' safety in schools is basically under control though it may have some kinds of discipline problems. In this kind of situation, the roles students' classroom lives played in forming their science performance need to be known. However, little quantitative research has been done in this area.

Theoretical Framework and Research Questions

As said above, countries that heavily relied on textbooks (TE) performing well in PISA science tests. Meanwhile, elements such as instruction (P), students' learning initiative (LI), and classroom life (CL) were also responsible for explaining students' science performance. So far, the majority of existing research in this domain was a "simple linear regression analysis" pattern because they only focused on a single influential factor of students' science performance. Findings that came from this kind of research are less convincing because they can be varied in the context where more than one factor influenced student's learning. However, up to now, little was known about the holistic mechanism of P, in combination with LI, TE, and CL, functioning on students' science performance. One of the methods to explore this holistic picture is multiple linear regression (MLR) analysis. But it also has some deficiencies in exploring the complex mechanism of the foregoing factors exerting on students' science performance. The reason for it is the foregoing factors are both taken as covariates in MLR analysis. Therefore, although these factors are taken into the statistic model simultaneously, they are not connected, but only connected to the outcome variable respectively. It is hard to believe that these factors are independent of each other. Being components of the instructional system, there must be some kinds of connections among them to make the system function well. A reasonable holistic picture requires not only including these influential factors in the model simultaneously but also showing their interactions. Therefore, a theoretical framework that provides a reasonable explanation for this kind of interaction should be found first.

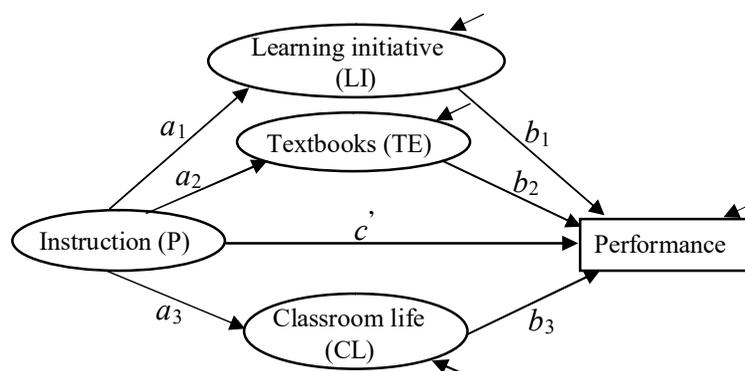
Since these factors focused on students' science learning, the instructional design model (ID model) can be a theoretical framework to integrate them. Instructional design is a paradigm that is related to encoding and decoding the messages (Gagne et al., 2005; Khalil & Elkhider, 2016; Ledford & Sleeman, 2000, p.13). It exposes that teaching and learning in classrooms is a process of message generation, flow, and assimilation. As far as assimilating the message is concerned, Reigeluth (1999) put forward that there are four conditions: "what is to be learned, the nature of learners, the learning environments and constraints." Constraints are something like money and time teachers owning to develop their instruction. As far as the condition "what is to be learned" is concerned, it relates to teachers' interaction with textbooks, thus could determine "what is to be learned" in a lesson. It is shown in figure 1 as a_2 path. For the condition "the nature of learners" is concerned, it means teachers' understanding of their students' characteristics and learning strategies, which in turn influence their instructional strategies, it is shown in figure 1 as a_1 path. In terms of the condition "the learning environment," it is built by the interactions of instruction and students' classroom life. It is shown in figure 1 as a_3 path. To achieve learning goals, these three paths all should point to the desired outcomes, which are b_1 path in figure 1. Based on systematic thinking, Dick et al. (2015) also put forward the components leading to the desired outcomes were "the instructor, learners, materials, instructional activities, delivery system, and learning and performance environments" (pp. 1-3). In this framework, the teacher, students, and materials are the static component in the instructional system. Teachers' adaptation of textbooks generates abundant learning materials. The interactions between the teacher and students generate instructional



activities and the delivery system. Since the learning environments are students and teachers combining into a collaborative group to solve problems (Dick et al., 2015, p.195), it is the interaction of instruction and students' classroom life, as well as the interaction of instruction and learning. For a performance environment, it is behaviors relating to students' learning initiatives (Dick et al., 2015, p.214), which is ignited by instruction. Therefore, Dick et al. (2015)'s framework also supports the statistical model shown in figure 1.

Figure 1

Statistical Diagram of the Mediation Effect of Instruction on Science Performance through Three Mediators



The key research questions were:

1. Do learning initiative (LI), textbooks (TE), and classroom life (CL) mediate the relation of instruction (P) and students' science performance?
2. Is gender a moderator of the following relations: instruction on learning initiative, learning initiative on performance, instruction on textbooks, textbooks on performance, instruction on classroom life, classroom life on performance, and instruction on performance?
3. Is gender a moderator of the three specific indirect effects and the total effect of instruction on the performance shown in figure 1?

Answers to these questions meant to depict a clear picture of science performance production in the school context.

Research Methodology

General Background

In this research, the school life questionnaire (SLQ) was developed with the aids of two officers who came from the city's teaching research office. This office is affiliated to the city's education bureau and responsible for the quality of the city's elementary and secondary education. The questionnaire selected items from several classroom observation protocols (Sawada et al., 2002; Weaver et al., 2005; Weiss et al., 2004), then modified some items based on the officers' opinions to fit the local environment. In developing the original SLQ, the criteria of specificity, clarity, and brevity were of utmost importance (Cowles & Nelson, 2015, p. 108; DeVellis, 2017, pp. 103-105; Dillman, 2009, p. 32; Fowler, 1995, p. 2). For specificity and clarity, no unfamiliar words or terms were used in the SLQ's item wordings. For the need for brevity, the words in the SLQ's item were as little as possible. The officers helped to arrange two focus group interviews in schools. It was a typical procedure in a pilot survey to test the quality of items (Cowles & Nelson, 2015, p. 128; Fowler, 1995, pp. 104-105). Based on low secondary school students' feedback, some of the item wordings were revised to make its meaning as clear as possible. As the original SLQ covered dozens of items, the interviewees also put forward their classmates would get bored to answer so many questions. Therefore, the length of the original SLQ needs to be reduced. Then, the questionnaire was applied to hundreds of lower secondary school students. According to the data gotten from the pre-test, the "corrected item-scale correlation" and "Cronbach's alpha if item deleted" values were used to evaluate the quality of an item. The item which its corrected item-scale correlation was smaller than .4 and its Cronbach's alpha if item



deleted was higher than the scale's Cronbach's alpha, was deleted. In this procedure, the key items were identified and constituted the final edition of SLQ. Then, 576 eighth-grade students completed the SLQ in December 2016. Confirmatory factor analysis (CFA), and exploratory structural equation modeling (ESEM) confirmed SLQ's four factors' structure and its composite reliability. Meanwhile, mediation analysis depicted a clear picture to explain the complex mechanism of LI, TE, and CL played in the relationship between P and science performance. However, some abnormalities were found hard to explain in the quantitative framework, such as why so many students who satisfied with their science teachers' instruction got poor performance in tests? It was a signal that depends on the quantitative method itself was not enough to expose the complex mechanism of students' science learning. For the needs of triangulation and complementarity (Hesse-Biber, 2010, pp. 3-4), three focus group interviews were conducted in December 2017. It allowed this research to make a thick description of students' opinions on their science learning initiative, classroom life, and textbook utilization.

Sample

This research selected participants from two public schools and one private school in Zhejiang province. These schools had good representativeness to the city's lower secondary schools. One public school is located in downtown. The other public school is located in the urban-rural conjunctive region. Both of them recruit students from the surrounding area. That private school is located downtown to recruit students from remote rural areas. Students were free to choose whether or not to participate in this research.

Procedures and Instruments

In consideration of students who may hesitate to tell the truth in the survey, the survey was conducted by student teachers. Lower secondary school students usually take student teachers as their elder sisters or elder brothers. They believe in their student teachers would not hand over their responses to in-service science teachers. In this light, they are more likely to tell their true feelings about school life, especially their opinions on the quality of their science teachers' instruction. As soon as they finished the questionnaire with a real name, the student teachers entered data in a spreadsheet and connected it to their science performance. In this semester, the four chapters in the science textbook are chemistry, geography, biology, and physics topics, respectively. Students will take an examination as soon as they finish one chapter. This educational system is named "the month test" and has decades of history in Zhejiang province. Typically, a student's scores on different topics are fluctuant, depending on one test could not determine the student's learning performance. The test papers, either used in the month tests or the final examination, are arranged by schools. It is hard to guarantee the test invariance of these different test papers. Since the three schools' quality of teaching decreases in turn, the downtown public school is the best, and the downtown private school is the worst, using the same test papers also would not help to this research. In that case, students with the same learning initiatives may achieve very different scores due to the school's quality of teaching is different. That may result in misunderstanding the mechanism of science performance production in the school context. Thus, students' science performance was evaluated by in-service science teachers according to their performance in the last six months' science tests with a five-grade system. Compared to the original score in one test, the five-grade rating was suitable to the SLQ and the different quality of teaching in schools. Eventually, 484 valid responses were obtained.

The research team conducted chi-square tests of independence with the questionnaire data to explore the dependence of variables, also explored the mechanism of antecedent variables functioning on students' performance by the structural equation modeling approach. Some abnormalities were hard to explain in a quantitative framework, such as why students who satisfied with their science teachers' instruction got poor performance in tests, why did so many students lose their confidence in learning science? Therefore, a qualitative approach would benefit this research, as it is complementary to the quantitative method. The next year of the foregoing survey had accomplished, two groups of student teachers came to the foregoing two public schools. In each group, there was a team leader who was not only responsible for managing the entire team to make sure its members cooperating well with school staffs, but also responsible for conducting focus group interviews. The interviews were oriented to find students' opinions on their teachers' instruction, textbook, classroom life, and their learning initiative. Using focus group interviews rather than individual interviews is due to its advantages in saving time and igniting participants to resonate with each other's experiences (Auerbach & Silverstein, 2003, p. 17). Every team leader



organized several discussions in her team to discuss the semi-structured interview questions developed by the research team. Then the research team held several discussions with two team leaders to reduce questions and to simplify wordings of some questions, and at the same time clarifying the interview goal. The research team purposively selected interviewees from the team leaders' classes rather than other student teachers' classes. In this case, interviewees were more likely to tell their opinions as they were familiar with the moderator. One team leader established two focus groups, each consisting of 14 eighth-grade students. One top student group in which its member's science performance was evaluated as level 4 or level 5, one low science proficiency student group in which its member's science performance was evaluated as level 1 through level 3. They were labeled as T_1 and L_1 groups respectively. Another team leader established one focus group consisting of 6 ninth-grade students. It was a low science proficiency student group and labeled as group L_2 . Every group had an equal number of male and female students. Two team leaders held three interviews. T_1 and L_1 groups' interviews lasted approximately forty minutes, which was one lecture time. The L_2 group's interview cost 30 minutes. In total, three group interviews were audiotaped and transcribed.

The SLQ has four subscales. One subscale is the independent variable P. The other three subscales in SLQ are mediation variables LI, TE, and CL. All of them are 5-point subscale, and each subscale has 5 observed indicators/items. Item wordings are detailed in table 1.

Table 1*The School Life Questionnaire (SLQ)*

Items	Wordings
Learning initiative (LI)	
LI1	I am interested in science.
LI2	I have confidence in learning science.
LI3	I always pay attention to what my teacher is saying.
LI4	I have the enthusiasm to answer my teacher's questions and attend science activities.
LI5	I use various strategies in learning science, such as preview, review, reflection, taking notes.
Instruction (P)	
P1	My science teachers did not do the whole talking. They allow us to discuss and explore in the class.
P2	My science teachers frequently invited me to answer their questions.
P3	I hold the feelings appreciated and encouraged by my science teacher.
P4	My science teacher's pedagogy draws my attention.
P5	My science teacher answers my questions on content knowledge that I don't understand.
Textbook (TE)	
TE1	The science textbook's writing style draws my attention.
TE2	The wordings, symbols, tables, and pictures in the science textbook are easy to read and understand.
TE3	The science textbook highlights key content knowledge.
TE4	The depth of the science textbook fits my learning ability.
TE5	The breadth of the science textbook fits my learning ability.
Classroom life (CL)	
CL1	I like my classmates.
CL2	My classmates keep discipline in the classroom.
CL3	My classmates cooperate well in learning science.
CL4	I played with my classmates happily after class.
CL5	My classmates and I feel proud of our class.

After the questionnaire survey, this research carried out a semi-structured interview later. It focused on dimensions specified by the SLQ and with intentions to dig deeper into students' perceptions of the influential factors



of their science learning in the school context and to find reasons behind their perceptions. In the interviews, the moderators asked students the following questions:

- Q1. Are you interested in science? What are the reasons that you like/dislike it?
- Q2. Do you have confidence in learning science? What is the reason for that?
- Q3. Are you willing to answer questions, do hands-on and minds-on activities, and participate in group discussion in science lessons? What is the reason for that?
- Q4. To what extent your science learning gains are the results of your teacher's instruction?
- Q5. How are your feelings about the roles your classmates played in your science learning?
- Q6. Are you interested in reading science textbooks? What is the reason for that?
- Q7. What do you think is the main reason of students' failure in science (question for top student group only)? What do you think is the main reason of students' success in learning science (question for low science proficiency student group only)?

Following the questioning, the students were free to discuss any issue which was related to their science learning.

Data Analysis

To verify the SLQ's 4 factors structure, also for the needs of calculating the SLQ subscales' reliability, the research carried out confirmatory factor analysis (CFA) through Mplus 7.4. As far as a scale which subscales are concerned, composite reliability (CR) is more suitable to be the estimator of a subscale's reliability than Cronbach's coefficient alpha (Bentler, 2009; Raykov, 2004; Raykov & Grayson, 2003; Sijtsma, 2009). In this case, the SLQ subscales' CR was calculated based on Mplus' output file of CFA. Meanwhile, the research conducted exploratory structural equation modeling (ESEM) to further examine the SLQ's factorial structure. Since the main purpose of this research was to explore whether there were mechanisms of P via LI, TE, and CL to science performance respectively, the multiple mediation analysis was performed to examine it. Besides, because it is of this research interests to explore whether there was any gender difference in these mechanisms, the multi-group structural equation model was constructed by Mplus 7.4 to answer it. Depending on the quantitative approach alone was not enough to explain the complex mechanisms of P, LI, TE, and CL exerting on science performance. Semi-structured interviews were administrated for the needs of complementing the quantitative approach. Then the interview transcripts were analyzed by a group of five researchers. They were the authors, two team leaders of student teachers. The interviews were analyzed according to the process of preparing, writing memos, coding, and presenting (Edmonds & Kennedy, 2017, pp. 321-331; Merriam & Tisdell, 2016, pp.196-199; Saldaña, 2013, pp. 41-52; Seidman, 2006, pp. 112-125). To ensure the validity of data analysis, the norms of avoiding prejudice, cross-check, and using memos were adhered to (Edmonds & Kennedy, 2017, p. 323; Merriam & Tisdell, 2016, p.208; Seidman, 2006, pp. 117-121). After reviewing the transcripts or audio files, the members of the research team wrote memos to record reflections independently, and group discussions were held to ensure no important messages were overlooked and the data were not contaminated by one of the researcher's biases or prejudice. The next steps were coding and presenting the reduced data according to the research interests. For example, in the interests of understanding students' classroom lives, the research team wrote memos such as the low proficiency students never complained about the discipline problems, whereas the top students complained about it lots. Then, they were coded as "uncooperative climate." The findings of the interviews were presented in the following sub-chapter of "Students' Introspection on Their Science Learning"

Research Results

Factorial Structure and Reliability of the SLQ

The CFA procedure was used to evaluate the extent to which the hypothesized factorial structure of the SLQ was true. For this purpose, the cut-off criteria suggested by the literature were adopted. They are $\chi^2/df < 3$ (Byrne et al., 1989; Marsh & Hocevar, 1985), standardized root mean square residual (SRMR) $< .08$ (Hu & Bentler, 1999) or SRMR $< .10$ (Marsh et al., 2005), root mean square error of approximation (RMSEA) $< .06$ (Hu & Bentler, 1999) or RMSEA $< .08$ (McDonald & Ho, 2002). The smaller the above-mentioned indices are, the better the model fits the data. As far as Tucker-Lewis index (TLI) and comparative fit index (CFI) are concerned, TLI $> .95$ and CFI $> .95$ mean a close



model fit (Hu & Bentler, 1999; Marsh et al., 2005), while $.90 < TLI < .95$ and $.90 < CFI < .95$ mean a fair model fit (Brown, 2014, p. 87; Wang & Wang, 2012, pp. 18-19).

To get standardized factor loadings, the STANDARDIZED statement was specified in the OUTPUT command in the Mplus program. SLQ subscales' CR was calculated through the standardized factor loadings. However, the CFA procedure restricted some factor loadings to zero, for example, P1 through P5, TE1 through TE5, and CL1 through CL5 not loaded onto factor LI, often brought poor-fitting CFA solutions (Brown, 2014, p. 193) and overestimate factor correlations (Asparouhov & Muthén, 2009). In this case, ESEM was also conducted to provide more information about the SLQ's factorial structure. Table 2 shows the SLQ's model fit indices.

Table 2
Model Fit Indices of the SLQ's 4 Factors Structure

Model	χ^2	df	χ^2/df	CFI	TLI	RMSEA		SRMR
						Value	CFit	
CFA	455.13	164	2.778	.924	.912	.061	.005	.048
ESEM	265.163	116	2.286	.961	.936	.052	.366	.026

Note. N=484. CFit: test of close fit.

As table 2 shows, the CFA procedure proves that the SLQ's 4 factors structure has an acceptable fit to the survey data, whereas the ESEM solution gets a better fit to data. The reason for saying so was ESEM achieved better fit indices, especially for CFit of RESEA was no longer significant ($p=.366 > .05$). Depending on the fit indices only was not enough to confirm the SLQ's 4 factors structure. It was also necessary to examine whether the primary loadings of the SLQ items were in accord with the prediction. Table 3 provides this kind of information.

Table 3
Items Factor Loadings and their z-test Results Based on ESEM

Items	Factor 1		Factor 2		Factor 3		Factor 4	
	Loading	z-score	Loading	z-score	Loading	z-score	Loading	z-score
LI1	.726	13.637***	.030	.651	.108	1.901	-.056	-1.458
LI2	.755	15.465***	.026	.644	-.044	-1.264	.020	.700
LI3	.612	11.027***	.016	.300	-.039	-.912	.017	.435
LI4	.526	7.472***	-.072	-1.046	.255	3.484***	.133	2.212*
LI5	.370	6.863***	.121	1.935	.051	1.011	-.025	-.586
P1	.000	.011	.377	5.758***	.351	5.962***	.015	.379
P2	.065	1.471	.030	.748	.817	12.973***	-.035	-1.084
P3	-.077	-1.617	.033	.667	.787	11.574***	.010	.309
P4	.180	2.835*	.454	5.847***	.346	4.826***	.088	1.708
P5	.203	3.184**	.303	3.985***	.261	3.724***	.145	2.621**
TE1	-.021	-.538	.691	10.903***	.083	1.538	-.040	-.983
TE2	-.002	-.080	.691	13.514***	-.023	-.607	-.005	-.151
TE3	-.051	-1.322	.779	14.766***	-.034	-.962	.013	.440
TE4	.122	2.517*	.604	10.606***	.035	.858	-.008	-.226
TE5	.138	2.654**	.555	9.154***	.025	.574	.045	1.108
CL1	.035	1.050	.192	2.989**	-.027	-.616	.556	10.621***
CL2	-.194	-2.780**	-.047	-.85	.176	2.644**	.643	10.862***



Items	Factor 1		Factor 2		Factor 3		Factor 4	
	Loading	z-score	Loading	z-score	Loading	z-score	Loading	z-score
CL3	.085	1.670	-.012	-.302	-.022	-.547	.853	15.607***
CL4	.078	1.438	.038	.589	.072	1.230	.653	10.992***
CL5	-.113	-1.937*	.100	1.423	-.020	-.428	.763	12.385***

Note. $N=484$. * $p < .05$, ** $p < .01$, *** $p < .001$; z-score is determined by dividing the unstandardized parameter estimate by its standard error (Brown, 2014, p. 128). Five largest loadings on every factor are in boldface.

As can be seen in table 3, items LI1 through LI5 have their largest loadings on factor 1, items TE1 through TE5 have their largest loadings on factor 2, and items CL1 through CL5 have their largest loadings on factor 4. Therefore, factors 1, 2, and 4 represented the latent dimensions of LI, TE, and CL, respectively. However, items P1, P4, and P5 had large and statistically significant loadings on TE. Since the five biggest loadings on factor 3 belonging to items P1 through P5 and the wordings of items P1 through P5 were not similar to items TE1 through TE5, these three cross-loadings were not a severe violation of the SLQ's factorial structure. The cause of their cross-loadings may root in teachers' instruction would improve students' perceptions of textbooks' quality. The subscales' CR were computed and shown in table 4.

Table 4

Subscales' Composite Reliability (N=484)

Subscales/Latent dimensions	Standardized loadings	Composite Reliability
		CR
LI	.750, .794, .621, .559, .535	.790
P	.670, .677, .602, .763, .663	.808
TE	.698, .739, .769, .752, .710	.854
CL	.692, .536, .765, .682, .658	.802

Table 4 shows all the latent dimensions have acceptable reliability (range of CR=.790 to .854). In summary, tables 2 through 4 support the SLQ's 4 factors structure. Since the SLQ was reliable and valid, it was suitable for using the SLQ survey data to explore the holistic mechanism of P in combination with LI, TE, and CL, functioning on students' science performance.

Mediation Roles of Learning Initiative, Textbook, and Classroom Life

As figure 1 shows, a_1 estimates the effect of P on LI, TE, and CL, respectively; b_1 estimates the effect of LI on science performance holding P, TE, and CL constant. The same is true for b_2 and b_3 . That will yield three specific indirect effects. For example, a specific indirect effect of P on science performance through LI is a_1b_1 . Therefore, the total indirect effect is $a_1b_1 + a_2b_2 + a_3b_3$; the total effect of P on science performance is $c = c' + a_1b_1 + a_2b_2 + a_3b_3$.

The result showed the parallel multiple mediation model depicted in figure 1 had acceptable fit indices. They were $\chi^2/df=2.87$, RMSEA=.062, SRMR=.050, CFI=.916, and TLI=.901. The overall tests of significance of path coefficients, indirect and direct effects are shown in table 5.



Table 5*Test of Significance of Path Coefficients, Specific Indirect Effects, Direct and Total Effect*

Parameter	Standardized path coefficient	Product of path coefficients	95% CI	
			LL	UL
Path coefficients	a_1	.676***	.603	.749
	a_2	.844***	.803	.894
	a_3	.585***	.505	.698
	b_1	.557***	.448	.704
	b_2	-.137	-.275	.036
	b_3	-.003	-.129	.114
Specific indirect effects	a_1b_1	.377***	.292	.509
	a_2b_2	-.116	-.233	.025
	a_3b_3	-.002	-.078	.063
Direct effect	c'	.003	-.150	.161
Total effect	c	.262***	.180	.359

Note. $N=484$. *** $p < .001$; CI=confidence interval; LL=lower limit, UL=upper limit.

According to table 5, all regression coefficients of a_i path are large ($a_1=.676$, $a_2=.844$, $a_3=.585$) and statistically significant ($p < .001$). Two students that differ by one unit on P were estimated to differ by .676 units in their learning initiative, .844 units in their perceptions of the textbook quality, .585 units in their perceptions of the classroom life, respectively. The path coefficient of b_1 was large (.557) and statistically significant ($p < .001$). It meant two students that differ by one unit on LI were estimated to differ by .557 units on their science performance holding TE, CL, and P constant. The latter two b_i paths were negative and insignificant. For path TE→performance, two students that differ by one unit on their perceptions of the textbook quality were estimated to differ by .137 units on their science performance holding LI, CL, and P constant, with those who were more satisfied with the textbook had worse performance. For path CL→performance, two students that differ by one unit on their perceptions of the classroom life were estimated to differ by .003 units on their science performance holding LI, TE, and P constant, with those who were more satisfied with the classroom life had worse performance.

This model had three specific indirect effects. The first indirect effect of P on performance was modeled through LI, estimated as .377 and statistically significant. Students who were more satisfied with their teachers' instruction getting more learning initiative than those less satisfied with their teachers' instruction ($a_1=.676$), which in turn was positively related to the promotion of their science performance ($b_1=.557$). The second indirect effect of P on performance was modeled through TE and statistically insignificant. Two students that differ by one unit on their perceptions of teacher's instruction were estimated to differ by .116 units on science performance, with those more satisfied with their teacher's instruction had worse performance (because b_2 is negative). The third indirect effect of P on performance was modeled through CL, estimated as -.002. Although two negative specific indirect effects (P→TE→performance and P→CL→performance), which came from the negative path coefficients of b_2 and b_3 , were statistically insignificant, they should not be ignored. The potential reasons for these negative effects will be discussed later in more detail.

The direct effect of P on performance was $c'=.003$. It estimates the amount by which two students that differ by one unit on P differ on their performance holding all mediator constant. Since it was trivial (.003) and statistically insignificant ($p=.967$), the effect of P on performance was completely mediated by the mediators. The total indirect effect was positive, it was .259. The total effect of P on performance was .262. The total indirect effect accounted for 98.8% of the effect of P on performance.

Testing Invariance of the Structural Path Coefficients and Indirect Effects across Female and Male Groups

For this research question 2, it required to test whether the three a_i path coefficients, three b_i path coefficients, and the direct effect of P on performance was invariance among students in the female group and male group.



For this research question 3, it required to test whether the indirect effect of P on performance through LI, TE, and CL respectively, as well as the total effect of P on the performance, were invariant among students in the female group and male group. If these path coefficients and effects are invariant between female and male groups, then gender is not a moderator.

A baseline structural equation model (SEM) for male and female groups respectively needs to be established first, so that the multi-group SEM modeling can be carried out later. Based on the information coming from modification indices (MI) for fixed parameters specified in the baseline model, there was only one error covariance having a MI larger than 10 in both male and female groups. In this case, the error covariance associated with items P3 and P2 which had the largest MI (15.800 in the male group, 22.458 in the female group) was set free estimated in baseline SEM models. It was found that the baseline models fit data well. The fit indices for the female group ($n=218$) were $\chi^2/df=1.281$, RMSEA=.036, SRMR=.054, CFI=.945, TLI=.936. The close fit test did not reject the null hypothesis of RMSEA \leq .05 in this group ($p=.965$). The fit indices for the male group ($n=266$) were $\chi^2/df=1.415$, RMSEA=.039, SRMR=.058, CFI=.927, TLI=.916. The close fit test also did not reject the null hypothesis of RMSEA \leq .05 ($p=.947$). Now that the baseline models fit data well, to test the invariance of structural path coefficients, indirect effects, and the total effect across two groups, an unrestricted SEM model using male and female samples simultaneously was established. Results of free estimated path coefficients, indirect effects, and total effect can be seen in table 6.

Table 6

Testing the Invariance of Structural Path Coefficients, Indirect Effects and Total Effect across Female and Male Groups

Parameter	Standardized estimates		Wald test ^a	
	Female group	Male group	χ^2	p
Structural path coefficient				
a_1	.737***	.616***	4.497*	.034
a_2	.881***	.837***	Not needed	
a_3	.620***	.603***	Not needed	
b_1	.676***	.509***	.609	.435
b_2	.078	-.189	Not needed	
b_3	.061	-.090	Not needed	
c	-.300	.113	2.182	.140
Indirect/Total effect				
a_1b_1	.498***	.313***	2.748	.097
a_2b_2	.069	-.158	1.503	.220
a_3b_3	.038	-.054	Not needed	
c	.305***	.214**	.901	.343

Note. $n_{\text{female}}=218$. $n_{\text{male}}=266$. * $p<.05$, ** $p<.01$, *** $p<.001$. a. $df=1$.

The unrestricted SEM model had an acceptable fit to data: $\chi^2/df=1.410$, RMSEA=.041, close-fit test $p=.971$, SRMR=.066, CFI=.918, TLI=.913. All a_i path coefficient, as well as b_1 , were positive and statistically significant in both male and female groups. They showed the same tendencies of antecedent variables on consequent variables (P \rightarrow LI, P \rightarrow TE, P \rightarrow CL, LI \rightarrow performance), but it seemed that the discrepancies of two path coefficients (i.e., a_1 , b_1) between male and female groups were large. The remaining three path coefficients showed different tendencies across groups. For example, the path coefficient of c , also known as the direct effect, was positive in the male group and negative in the female group. The discrepancy of c across groups was also large. Then three restricted SEM models were established to test the invariance of a_1 , b_1 , and c across groups. The MODEL TEST command in the Mplus program was used to provide a Wald χ^2 testing information with $df=1$. It can be seen in table 6, the path coefficients of a_1 in different groups are not identical to each other ($\chi^2=4.497$, $p=0.034$). It means P was more effective in igniting girls' LI than boys. In other words, the direct effect of P on LI was moderated by gender. The remaining two Wald tests on the invariance of b_1 and c across groups did not reject the null hypothesis: $\chi^2=.609$ ($p=.435$) for b_1



and $\chi^2=2.182$ ($p=.140$) for c' . Since b_1 and c' were invariant across groups, it was not necessary to test the invariance of a_2 , a_3 , b_2 , and b_3 across groups, as they had a smaller discrepancy across male and female groups than b_1 and c' .

As far as the indirect effects were concerned, the Wald test was used to examine the invariance of the specific indirect effects that P on performance through LI, as well as P on performance through TE. To be specific, the MODEL CONSTRAINT command set the indirect effect, accompanied by the MODEL TEST command to provide a Wald test. Both the specific effect of $P \rightarrow LI \rightarrow \text{performance}$ ($\chi^2=2.748$, $p=.097$) and the specific effect of $P \rightarrow TE \rightarrow \text{performance}$ ($\chi^2=1.503$, $p=.220$) were insignificant. There was not necessary to test the invariance of the remaining specific indirect effect ($P \rightarrow CL \rightarrow \text{performance}$) before accepting the null hypothesis, as this indirect effect had the smallest discrepancy between male and female groups. Testing for the invariance of the total effect of P on performance also resulted in accepting the null hypothesis ($\chi^2=.901$, $p=.343$). Although it was not the main purpose of this section, it should not be overlooked that the total effect of P on performance in both groups was statistically significant.

Students' Introspection on Their Science Learning

The negative path coefficients, such as b_2 , b_3 in male students baseline SEM model and c' in female students baseline SEM model, as well as b_2 , b_3 in the parallel mediation model, could be explained by the SLQ survey data. A negative estimate reflected the inconsistency between the antecedent variable and the consequent variable. For example, it was found that only 87 female students (39.9%, $n=218$) whose assessment levels on the factor P were suitable to their levels of performance. The others thus contributed to explaining the cause of the negative path coefficient of c' . However, depending on this technical analysis alone was not enough to answer this research concerns. The following issues are also of this research interests. What characteristics deeply hide behind classroom life and textbooks to impede them from acting as positive influential factors to students' science performance? What are the intrinsic defects in direct instruction to weaken its strength on students' learning outcomes? Answers to these questions are very important because it could complement some missing pieces to the holistic science performance picture. The following statements covered some relevant information obtained from interviews.

Q1. All students were interested in biology because it had close relations with the human body and daily life. In the meantime, all students were not interested in the difficult part of the science course. Low proficiency students were also interested in observing or doing experiments. Students in group L_2 were interested in chemistry because their teacher did many chemistry experiments. They also indicated that physics was the most difficult part. Students in group L_1 often got nervous in science lessons because the science teacher was strict with them.

Q2. All top students had confidence in learning science. They said cleverness was the source of their confidence. Almost all students in group L_2 and 5 students (35.7%, $n=14$) in group L_1 lost their confidence in learning science. 9 students in group L_1 (64.3%, $n=14$) had a little confidence in science because they could understand the easy part of the science course.

Q3. For the top students, the average time they stayed focus on science tasks was three-quarters of a class. Three-quarters of them were not willing to answer questions as they may make mistakes. For the low proficiency students, the minimum time they focused on science tasks was 5 minutes, with the whole lecture lasted 40 minutes. They were bored in the lecture that the teacher did most of the talking and rarely did experiments. In this context, even top students complained that they felt exhausted when the lesson was over. However, when there was an experiment, the low proficiency students were willing to engage in it, even having discussions with their teachers, which they rarely did in other contexts. Compared to the top students, the low proficiency students were more reluctant to answer questions for the same reason.

The first three interview questions focused on students' learning initiatives. In sum, no matter what levels they got in science performance, they were all fond of biology and not interested in the difficult part of the science course. Their confidence in learning science was hurt badly by the difficult part of the science course. Cleverness was the only factor put forward by the top students to account for their confidence in learning science. Meanwhile, the low proficiency students were absent from the science lesson in most of the time unless there was an experiment.

Q4. Top students appreciated their teacher's instruction lots. They acknowledged that if they had studied by themselves, they would not have gained much in science. Students in group L_1 agreed that the extent to which the science teacher contributing to their studying was about half of all gains. Students in group L_2 considered that the science teacher gave little help to their learning. Though he taught well, the previous science teacher had taught bad, thus they did not prepare well for the current studying.

Q5. Students in group L_1 took top students as obstacles to their science learning. They felt upset for not as clever as the top students. Students in group L_2 hold the same attitudes toward the top students, they complained top students for causing the pace of the science lesson too fast. Moreover, the top students did not help them in learning science. Top students appreciated each other for their endeavor to construct an atmosphere of studying hard. They complained that the low



proficiency students chatting in the lecture always broke their thoughts. Besides, some of their classmates were chasing and shouting during the break time, which impeded them to write the homework.

Q6. Students in group T_1 and half of the students in group L_1 were interested in reading the science textbook. The biological knowledge, experiments, and pictures in the textbook attracted their attention. Yet all students in group L_2 reported that they were not interested in reading the textbook. Students in group T_1 suggested to mark key SMK in the text and connect SMK to real life to make it more comprehensible to them. Besides, they thought electromagnetic induction was too abstract to understand. No matter how hard they had worked on this topic, they cannot resolve problems related to electromagnetic induction. Students in group L_1 put forward that textbooks should have more experiments and pictures to draw their attention.

Q7. The top students summarized three main reasons to account for the low proficiency students' failure in a science course, i.e., not smart, not following teachers' instructions, and copying homework. By contrast, the main reasons proposed by the low proficiency students to explain the top students' success in science were learning well in previous science courses, studying tirelessly, and accommodating to teacher's pedagogy.

It can be seen from students' responses to Q4 and Q7, students acknowledged that the science teacher contributed greatly to the top students' learning. By contrast, the instruction did not have much influence on the low proficiency students' learning. They were not suited to the teacher's pedagogy, and the pace of the science lesson was too fast to follow. Students' responses to Q5 provided useful information on the CL dimension. It seemed that the top students and the low proficiency students did not support each other in learning science, whereas acting as an obstacle to impede another group's students to improve their gains. The top students complained a lot about the discipline issues in the classroom, while the low proficiency students said nothing about it. Students' responses to Q6 provided useful information on the TE dimension. The top students more frequently read the textbook than the low proficiency students. Advice proposed by the top students related to the textbook was mainly about making the text more convenient for them to understand the science content, whereas the low proficiency students required more pictures and experiments in the text to ignite their interests in science.

Discussion

For this research question 1, it turns out that LI is the only significant mediator. Then the focus group interviews were conducted to find the underlying causes for this mechanism. The top students reported that the science teacher gave them lots of help in learning science. Meanwhile, the low proficiency students put forward "adapting to the science teachers' instruction" as one main reason to account for the top students' success. Even the low proficiency students could stay focus on science lessons where there was an experiment or having relations with real life. Therefore, it seems the effect of P on performance would not come to fruition unless the instruction ignites students' learning initiative (LI). That is the primary mechanism of direct instruction affecting students' science learning. Direct instruction may not be necessary an "inferior pedagogy," so long as it could ignite students' learning initiatives.

Some strategies suggested by previous research again emphasized by students in this research, such as relating science content with real-life (Hulleman & Harackiewicz, 2009) and accommodating teachers' instruction to students' learning levels (Kremer et al., 2013). In students' viewpoints, relating science content with real life as well as teachers' experimental demonstrations will make SMK more comprehensible. Students also complained that they could never learn some difficult SMK, such as electromagnetic induction. Now that this kind of SMK is beyond lower secondary school students' learning ability, it may be reasonable to transfer them into the upper secondary school science course. Moreover, although previous research had found that students learned less with teachers talking more (Setati et al., 2002), this research found this argument may not be suitable for clever students. However, much attention should be paid to low proficiency students. They do not adapt to the content-heavy lessons where the pace is too fast to follow. As they do not understand science contents well, lots of them are also not interested in reading textbooks.

In practice, some people argue that students' success owes to their teachers little, as opposed to this opinion, others believe instruction is the main reason account for students' success. The research could not agree with these opinions. The former only pays attention to the insignificant direct effect of c' path. It may underestimate teachers' efficiency in prompting students' learning initiatives that, in turn, improve students' science performance. The latter could not differentiate the specific indirect effect of a_1b_1 from the direct effect of c' ; in other words, a_1b_1 is misunderstood as c' . In that case, it may underestimate the importance of adapting instruction to students' needs to ignite students' learning initiatives. Therefore, both of them are contrary to the mechanism that the effect of P on performance is completely mediated by LI.



P has significant direct effects on TE and CL. It means P actually communicates with TE and CL, in this light, it is reasonable for the mediation analysis to adopt ID models as its theoretic framework. However, the structural path coefficients of b_2 and b_3 are insignificant, which also causes insignificant indirect effects of a_2b_2 and a_3b_3 . It might indicate that the science teachers' instruction can compromise to TE and CL rather than producing fundamental alterations in science content presentation and classroom life. In compromising with TE and CL, P could improve students' satisfaction with the textbook and the classroom life to some extent, and this leads to the significant path coefficients of a_2 and a_3 . Since the intrinsic defects of TE and CL have not been fundamentally changed, the effects of TE on performance, and CL on performance, are destroyed. As far as the textbook was concerned, the inherent flaws included not connecting to real life, lacking experiments and interesting pictures, having some difficult content knowledge, and so on. As far as the classroom life was concerned, the low proficiency students complained that the top students did not give them practical help in learning science, the top students complained that chaos in classrooms always broke their thoughts. Therefore, students' uncooperative attitude may cause an insignificant direct effect of CL on performance.

For this research questions 2 and 3, only the relation of P on LI has gender difference. To be specific, the instruction is more effective in promoting girls' learning initiatives because the Wald test on the invariance of a_1 across groups rejected the null hypothesis. But what caused this difference between groups? Previous studies had found that the secular humanist society usually viewed women as responsible for housekeeping (Normile, 2006; Saujani, 2017). This gender stereotype may explain girls' tendencies of doing things quietly and gently. Compared to boys, they are more likely to cooperate with their science teachers and therefore have a more positive reaction to the science teachers' instruction.

Conclusions

Through mediation analysis, this research shows instruction has a significant influence on students' science performance. The total effect of instruction on performance in both male and female groups are statistically significant. Meanwhile, the impact of instruction on performance is completely mediated by students' learning initiatives. It exposes that even in the context of direct instruction, teachers cannot play as authorities, but should be facilitators. The higher instruction prompting students' learning initiatives, the more it gains in desired outcomes. Both boys and girls can succeed in the conditions that they have learning initiative. In terms of gender difference, it was found by multi-group SEM, although girls are more sensitive to teachers' pedagogy than boys, the strength of instruction on performance is not significantly higher than boys.

This research emphasizes the impact of instruction on performance can be improved by the visualization of science contents. Students required more pictures and experiments in textbooks, as well as more experiment demonstrations in science lessons. Students believe in what they see in their eyes. This kind of materials and classroom activities build an effective delivery system to make science concepts comprehensible. In this case, students can stay focus on science tasks.

This research finds the impact of instruction on performance is impaired by students' uncooperative classroom life. Teachers' beliefs about teaching shaped the classroom climate. In an exam-centric education system and content-heavy classrooms, discipline problems exist to some degree. It is not only a signal of students' revolt to tedious instruction but also a signal of students' dissatisfactions that the top students do not help them to learn science. In this classroom context, students' classroom life cannot play as a channel of communication between instruction and performance. This research also finds the impact of instruction on performance is impaired by some defects of textbooks. Those defects account for students losing their interests and confidence in learning science. Further research can explore the mechanisms of instruction exerting on science performance in an inquiry teaching context or explores whether family factors (e.g., parents' support) moderate these mechanisms.

Acknowledgements

The authors would like to thank Ling-hui Liang and Chang-guan Li, the city's educational officials, for their help with developing the school life questionnaire, as well as the support of the National Natural Science Foundation of China [Grant number 11975012].



References

- Akinoglu, O., & Yasar, Z. (2007). The effects of note taking in science education through the mind mapping technique on students' attitudes, academic achievement and concept learning. *Journal of Baltic Science Education*, 6(3), 34-43. <http://oaji.net/articles/2014/987-1404288606.pdf>
- Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science*, 333(6045), 1034-1037. <https://doi.org/10.1126/science.1207998>
- Aşkın, O. E., & Öz, E. (2020). Cross-national comparisons of students' science success based on gender variability: Evidence from TIMSS. *Journal of Baltic Science Education*, 19(2), 186-200. <https://doi.org/10.33225/jbse/20.19.186>
- Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, 16(3), 397-438. <https://doi.org/10.1080/10705510903008204>
- Auerbach, C. F., & Silverstein, L. B. (2003). *Qualitative data: An introduction to coding and analysis*. New York University Press.
- Banilower, E. R., Smith, P. S., Weiss, I. R., Malzahn, K. A., Campbell, K. M., & Weis, A. M. (2013). *Report of the 2012 national survey of science and mathematics education*. Horizon Research. <http://www.horizon-research.com/2012nssme/wp-content/uploads/2013/02/2012-NSSME-Full-Report1.pdf>
- Bao, L., Cai, T., Koenig, K., Fang, K., Han, J., Wang, J., Liu, Q., Ding, L., Cui, L. L., Luo, Y., Wang, Y. F., Li, L. M., & Wu, N. L. (2009). Learning and scientific reasoning. *Science*, 323(5914), 586-587. <https://doi.org/10.1126/science.1167740>
- Beaman, L., Duflo, E., Pande, R., & Topalova, P. (2012). Female leadership raises aspirations and educational attainment for girls: A policy experiment in India. *Science*, 335(6068), 582-586. <https://doi.org/10.1126/science.1212382>
- Bentler P. M. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika*, 74(1), 137-143. <https://doi.org/10.1007/s11336-008-9100-1>
- Brophy, J. (2006). History of research on classroom management. In C. M. Evertson & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 17-43). Lawrence Erlbaum Associates Publishers.
- Brown, T. A. (2014). *Confirmatory factor analysis for applied research*. Guilford Publications.
- Burušić, J., Šimunović, M., & Šakić, M. (2019). Technology-based activities at home and STEM school achievement: The moderating effects of student gender and parental education. *Research in Science & Technological Education*, 37(4), 1-22. <https://doi.org/10.1080/02635143.2019.1646717>
- Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, 105(3), 456-466. <https://doi.org/10.1037/0033-2909.105.3.456>
- Change the equation: Improving science and mathematics education in the US [Editorial]. (2011). *Nature Cell Biology*, 13(8), 875. <https://doi.org/10.1038/ncb2318>
- Cobern, W. W., Schuster, D., Adams, B., Applegate, B., Skjold, B., Undreiu, A., Loving, C. C., & Gobert, J. D. (2010). Experimental comparison of inquiry and direct instruction in science. *Research in Science & Technological Education*, 28(1), 81-96. <https://doi.org/10.1080/02635140903513599>
- Coll, R. K., Dahsah, C., & Faikhamta, C. (2010). The influence of educational context on science learning: A cross-national analysis of PISA. *Research in Science & Technological Education*, 28(1), 3-24. <https://doi.org/10.1080/02635140903513532>
- Cowles, E. L., & Nelson, E. (2015). *An introduction to survey research*. Business Expert Press.
- DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). SAGE Publications.
- Dick, W., Carey, L., & Carey, J. O. (2015). *The systematic design of instruction* (8th ed.). Pearson.
- Dillman, D. A. (2009). *Internet, mail and mixed-mode surveys: The tailored design method*. John Wiley & Sons.
- Djigic, G., & Stojiljkovic, S. (2011). Classroom management styles, classroom climate and school achievement. *Procedia-Social and Behavioral Sciences*, 29, 819-828. <https://doi.org/10.1016/j.sbspro.2011.11.310>
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4-58. <https://doi.org/10.1177/1529100612453266>
- Edmonds, W. A., & Kennedy, T. D. (2017). *An applied guide to research designs: quantitative, qualitative, and mixed methods* (3rd ed.). SAGE Publications.
- Fine, C., & Elgar, M. A. (2017). Promiscuous men, chaste women and other gender myths. *Scientific American*, 317(3), 32-37. <https://doi.org/10.1038/scientificamerican0917-32>
- Fowler, F. J., Jr. (1995). *Improving survey questions: Design and evaluation*. SAGE Publications.
- Fraser, B. J. (1987). Classroom learning environments and effective schooling. *Professional School Psychology*, 2(1), 25-41. <https://doi.org/10.1037/h0090526>
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410-8415. <https://doi.org/10.1073/pnas.1319030111>
- Gagne, R. M., Wager, W. W., Golas, K. C., Keller, J. M., & Russell, J. D. (2005). Principles of instructional design. *Performance Improvement*, 44(2), 44-46. <https://doi.org/10.1002/pfi.4140440211>
- Gieve, S., & Miller I.K. (2006). What do we mean by 'quality of classroom life'? In S. Gieve & I. K. Miller (eds.), *Understanding the language classroom* (pp. 18-46). Palgrave Macmillan. https://doi.org/10.1057/9780230523166_3
- Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Culture, gender, and math. *Science*, 320(5880), 1164-1165. <https://doi.org/10.1126/science.1154094>



- Gupta, M. D. (2017). Return of the missing daughters. *Scientific American*, 317(3), 80-85. <https://doi.org/10.1038/scientificamerican0917-80>
- Hanbay, O. (2015). Student survey on the textbook 'DEUTSCH IST SPITZE! A1. 1'. *International Journal of Language Academy*, 3(1), 319-324. <https://doi.org/10.18033/ijla.208>
- Hesse-Biber, S. N. (2010). *Mixed methods research: Merging theory with practice*. The Guilford Press.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55. <http://doi.org/10.1080/10705519909540118>
- Hulleman, C. S., & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. *Science*, 326(5958), 1410-1412. <http://doi.org/10.1126/science.1177067>
- Johnson, B., Sullivan, A. M., & Williams, D. (2009). A one-eyed look at classroom life: Using new technologies to enrich classroom-based research. *Issues in Educational Research*, 19(1), 34-47.
- Khalil, M. K., & Elkhider, I. A. (2016). Applying learning theories and instructional design models for effective instruction. *Advances in Physiology Education*, 40(2), 147-156. <http://doi.org/10.1152/advan.00138.2015>
- Kremer, M., Brannen, C., & Glennerster, R. (2013). The challenge of education and learning in the developing world. *Science*, 340(6130), 297-300. <http://doi.org/10.1126/science.1235350>
- Ledford, B. R., & Sleeman, P. J. (2000). *Instructional design: A primer*. Information Age Publishing.
- Lim, W. K. (2010). Asian education must change to promote innovative thinking. *Nature*, 465(7295), 157. <http://doi.org/10.1038/465157a>
- Machin, S., & Pekkarinen, T. (2008). Global sex differences in test score variability. *Science*, 322(5906), 1331-1332. <http://doi.org/10.1126/science.1162573>
- Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit in structural equation models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Multivariate applications book series. Contemporary psychometrics: A festschrift for Roderick P. McDonald* (pp. 275-340). Lawrence Erlbaum Associates Publishers.
- Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First- and higher order factor models and their invariance across groups. *Psychological Bulletin*, 97(3), 562-582. <https://doi.org/10.1037/0033-2909.97.3.562>
- Matsumura, L. C., Slater, S. C., & Crosson, A. (2008). Classroom climate, rigorous instruction and curriculum, and students' interactions in urban middle schools. *The Elementary School Journal*, 108(4), 293-312. <https://doi.org/10.1086/528973>
- McDonald, R. P., & Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, 7(1), 64-82. <https://doi.org/10.1037/1082-989X.7.1.64>
- Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation (4th ed.)*. John Wiley & Sons.
- Normile, D. (2006). Getting women scientists back on the career track in Japan. *Science*, 311(5765), 1235-1236. <https://doi.org/10.1126/science.311.5765.1235>
- Oates, T. (2014). *Why textbooks count: A policy paper*. University of Cambridge. <https://www.cambridgeassessment.org.uk/Images/181744-why-textbooks-count-tim-oates.pdf>
- Organisation for Economic Co-Operation Development. (2016). *PISA 2015 results in focus*. OECD. <https://doi.org/10.1787/aa9237e6-en>
- Organisation for Economic Co-Operation Development. (2019a). *PISA 2018 results (volume I): What students know and can do*. OECD. <https://doi.org/10.1787/5f07c754-en>
- Organisation for Economic Co-Operation Development. (2019b). *PISA 2018 results (volume II): Where all students can succeed*. OECD. <https://doi.org/10.1787/b5fd1b8f-en>
- Pierce, C. (1994). Importance of classroom climate for at-risk learners. *The Journal of Educational Research*, 88(1), 37-42. <https://doi.org/10.1080/00220671.1994.9944832>
- Raykov, T. (2004). Behavioral scale reliability and measurement invariance evaluation using latent variable modeling. *Behavior Therapy*, 35(2), 299-331. [https://doi.org/10.1016/S0005-7894\(04\)80041-8](https://doi.org/10.1016/S0005-7894(04)80041-8)
- Raykov, T., & Grayson, D. A. (2003). A test for change of composite reliability in scale development. *Multivariate Behavioral Research*, 38(2), 143-159. https://doi.org/10.1207/S15327906MBR3802_1
- Reichenberg, M. (2016). Explaining teachers' use of textbooks. *Journal of Educational Media, Memory, and Society*, 8(2), 145-159. <https://doi.org/10.3167/jemms.2016.080208>
- Reigeluth, C. M. (1999). What is instructional-design theory and how is it changing? In C. M. Reigeluth (Ed.), *Instructional-design theories and models: A new paradigm of instructional theory* (Vol.2, pp. 5-29). Lawrence Erlbaum Associates Publishers.
- Robinson, S. (2011). Student use of a free online textbook. *Academy of Educational Leadership Journal*, 15(3), 1-10. <https://www.questia.com/library/journal/1G1-263157454/student-use-of-a-free-online-textbook>
- Rushton, S. P. (2000). Student teacher efficacy in inner-city schools. *The Urban Review*, 32(4), 365-383. <https://doi.org/10.1023/A:1026459809392>
- Sachs, J., & Mellor, L. (2005). 'Child panic', risk and child protection: An examination of policies from New South Wales and Queensland. *Journal of Education Policy*, 20(2), 125-140. <https://doi.org/10.1080/0268093052000341368>
- Saldaña, J. (2013). *The coding manual for qualitative researchers (2nd ed.)*. SAGE Publications.
- Saujani, R. (2017). Girl code. *Scientific American*, 317(3), 66-69. <https://doi.org/10.1038/scientificamerican0917-66>
- Sawada, D., Piburn, M. D., Judson, E., Turley, J., Falconer, K., Benford, R., & Bloom, I. (2002). Measuring reform practices in science and mathematics classrooms: The reformed teaching observation protocol. *School Science and Mathematics*, 102(6), 245-253. <https://doi.org/10.1111/j.1949-8594.2002.tb17883.x>
- Science education reforms in the UK [Editorial]. (2012). *Nature Cell Biology*, 14(10), 977. <https://doi.org/10.1038/ncb2601>
- Seidman, I. E. (2006). *Interviewing as qualitative research: A guide for researchers in education and the social sciences (3rd ed.)*. Teachers College Press.



- Setati, M., Adler, J., Reed, Y., & Bapoo, A. (2002). Incomplete journeys: Code-switching and other language practices in mathematics, science and English language classrooms in South Africa. *Language and Education*, 16(2), 128-149. <https://doi.org/10.1080/09500780208666824>
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74(1), 107-120. <https://doi.org/10.1007/s11336-008-9101-0>
- Sotáková, I., Ganajová, M., & Babinčáková, M. (2020). Inquiry-based science education as a revision strategy. *Journal of Baltic Science Education*, 19(3), 499-513. <https://doi.org/10.33225/jbse/20.19.499>
- The science of education reform [Editorial]. (2006). *Nature Neuroscience*, 9(11), 1345. <https://doi.org/10.1038/nn1106-1345>
- Wang, J. C., & Wang, X. Q. (2012). *Structural equation modeling: Applications using Mplus*. Higher Education Press; John Wiley & Sons Commercial Service (Beijing).
- Weaver, D., Dick, T., Higgins, K., Marrongelle, K., Foreman, L., & Miller, N. (2005). *OMLI classroom observation protocol*. RMC Research Corporation. <https://scaleupstem.ncsu.edu/OMLIClassroomObservationProtocols.pdf>
- Webb, P. (2010). Science education and literacy: Imperatives for the developed and developing world. *Science*, 328(5977), 448-450. <https://doi.org/10.1126/science.1182596>
- Weiss, I. R., Banilower, E. R., & Shimkus, E. S. (2004). *Local systemic change through teacher enhancement: Year nine cross-site report*. Horizon Research. <http://www.horizon-research.com/horizonresearchwp/wp-content/uploads/2013/04/cross-site03.pdf>
- Wilkens, H. J. (2011). Textbook approval systems and the program for international assessment (PISA) results: A preliminary analysis. *IARTEM e-Journal*, 4(2), 63-74. http://biriwa.com/iartem/ejournal/volume4.2/papers/Paper4_Wilkens_ApprovalSystemsandPISA_Final_Vol4No2.pdf

Received: September 14, 2020

Accepted: January 08, 2021

Cite as: Jiang, T., Chen, J.-G., & Wu, Y.-Y. (2021). Impact of instruction on science performance: Learning initiative as a mediator and gender as a limited moderator. *Journal of Baltic Science Education*, 20(1), 50-66. <https://doi.org/10.33225/jbse/21.20.50>

Tao Jiang
(Corresponding author)

PhD, Professor, Department of Physics, School of Electronics & Information Engineering, Taizhou University, 1139 Shifu avenue, Taizhou 318000, China.
E-mail: hopejt@163.com
ORCID: <https://orcid.org/0000-0001-8330-3995>

Ji-gen Chen

PhD, Professor, Department of Materials Engineering, School of Pharmaceutical and Materials Engineering, Taizhou University, Taizhou 318000, China.
E-mail: kiddchen@126.com
ORCID: <https://orcid.org/0000-0001-7580-3869>

Yin-yin Wu

PhD, Associate Professor, Department of Primary Education, College of Teacher Education, Taizhou University, 605 Dongfang avenue, Linhai 317000, China.
E-mail: wuyinyinbox@126.com

