JOURNAL OF BALTIC SCIENCE EDUCATION

ISSN 1648-3898 /Print/
ISSN 2538-7138 /Online/

CROSS-NATIONAL COMPARISONS OF STUDENTS' SCIENCE SUCCESS BASED ON GENDER VARIABILITY: EVIDENCE FROM TIMSS

Öyküm Esra Aşkın,
Ersoy Öz

Abstract. *Identifying key determinants of success plays an important role in efforts of assessing the current state of educational practices and implementing effective actions to improve academic performance. Given how important is accurate identification of these determinants for valid comparisons, the aim of this research was to present extensive comparative results on 8th grade students' science success for 39 countries. To this end, a total number of 229,493 students' information obtained from the last cycle of Trends in International Mathematics and Science Study was handled. Unlike previous studies in which success has been tried to be explained with some factors, this research extended the literature by investigating the effective factors as well as gender-related variability on science success with a holistic view. Using heterogeneous choice models, unobserved heterogeneity caused by gender difference was tested and controlled. The findings showed that gender variability was evident for 10 countries. Furthermore, factors associated with students' future educational goals, home educational resources and confidence were found as the common determinators of science success in all participant countries. Another notable finding was all factors, except time spent on doing homework, tended to differentiate the success status between girls and boys of 22 countries.*

Keywords: *cross-country comparison, gender variability, heterogeneous choice models, science success.*

Öyküm Esra Aşkın, Ersoy Öz
Yildiz Technical University, Turkey

Introduction

Identifying the underlying determinants (factors) of students' academic success is one of the topics that attract the most attention of educators and educational policy makers. An increasing number of studies have been conducted to put forth the relationship between academic success and effective factors in recent years. As stated by Martínez Abad and Chaparro Caso López (2017), researchers are faced with a complex task due to the diversity of factors and their relations. The successful execution of this task in terms of researchers may require, firstly, to work with reliable data sets, and secondly, to apply appropriate modelling strategies according to the structure of data.

Trends in International Mathematics and Science Study (TIMSS) is one of the largest international assessment studies which can serve as a source of reliable, comparable and representative data on different content areas for 4th and 8th grade students. TIMSS was first conducted in 1995 with more than 40 countries and different aspects of students' success have been evaluated every 4 years since then. Insights from TIMSS not only provide a better understanding of educational systems of participating countries (Mullis et al., 2012), but also give researchers the opportunity to make comparison of students' success between countries. Information coming from TIMSS has been seen as a helpful guide in determining the factors that significantly affect the students' success and has been used in a number of important studies that may be regarded as pioneers in this respect (e.g. Bouhlila, 2011; Neuschmidt, Barht, & Hastedt, 2008; Ramirez, 2006; Schreiber, 2002; Wößmann, 2003; Wößmann, 2005a). Since findings from models being estimated can be taken as a reference in shaping educational policies and restructuring of educational programs, choosing the appropriate modelling strategy plays a crucial role. In most of studies based on TIMSS results, students' success has been considered as a dependent variable and modelled with different interrelated aspects of students' success. Despite the fact that statistical modelling theory requires considering all effective factors on dependent variables, it is often impossible to fulfil this requirement due to the unobservable (i.e. omitted) characteristics correlated with observable factors, referred to as unobserved heterogeneity (Mood, 2009). Researchers have

https://doi.org/10.33225/jbse/20.19.186

increasingly tackled the effects of different dimensions of heterogeneity including differences in individual-based, school-based and time-based characteristics on students' success (Thomson, 2008; Wößmann, 2005b). The most commonly applied technique which implicitly accounts for the problem of heterogeneity is building nonlinear models (e.g. logit, probit, Poisson) with interaction terms. Nevertheless, the conclusions on estimated coefficients of the interaction terms have been regarded as problematic in traditional models such as logit (Tutz, 2018). Since ignoring unobserved heterogeneity may cause invalid and misleading results, different methods have been pursued to overcome the problem of bias in coefficients that is caused by changes in residual variances across compared groups (see, for example, Allison, 1999; Karlson, Holm, & Breen, 2012; Mood, 2009; Mustillo, Landerman, & Land, 2012; Rohwer, 2015; Williams, 2009). The usage of these methods has become much more recommended for quantitative research in recent years (Mustillo, Lizardo, & McVeigh, 2018). However, limited studies in educational literature are available on testing and controlling the unobserved heterogeneity. With this research, an important limitation of earlier studies in which efforts on explaining the relationship between success and its determinants has been done without considering the unobserved heterogeneity was eliminated. Specifically, gender variability was emphasized, and models were built for the purpose of making comparisons of effective factors on students' science success between TIMSS participating countries. Moreover, how the effects of factors on students' science success differ with respect to gender groups was searched.

Gender Heterogeneity in Success

Gender equity is considered as one of the major concerns for democratic societies (Baye & Monseur, 2016). Since gender differences in education plays the main role for careers in STEM fields (i.e. science, technology, engineering, and mathematics) (Hyde & Linn, 1988), closing the success gap between girls and boys in different content areas has been an important goal for developing countries' educational policies over decades. More specifically, for science, TIMSS reports published for each testing cycle shed light on gender gap at an international level (see, for example, Mullis et al., 2000; Mullis et al., 2004a; Mullis et al., 2004b; Mullis et al., 2016). Based on these reports, it is possible not only to examine which countries have gender differences in students' science scores, but also to capture the changes of gender gap in scores over the years. According to the current report (Mullis et al., 2016), significant differences in 8th graders' science scores are evident in 19 countries. The number of countries which girls are more successful compared to boys is remarkable. Girls have significantly higher scores in Bahrain, Botswana, Jordan, Kuwait, Lebanon, Malaysia, Malta, Morocco, Oman, Qatar, Saudi Arabia, Thailand, Turkey and United Arab Emirates, while boys are more successful than girls in Chile, Hong Kong SAR, Hungary, Italy and United States. It is worth to mention that these results are only based on scores. In other words, reported gender differences reflect the overall science scores of students without considering the effects of other factors on success.

It has been well documented that success varies within the group of girls and boys. According to the Maccoby and Jacklin (1974), boys have larger variance in some scores of mental tests (i.e. mathematical and spatial abilities) when compared to girls. Similar findings are also evident which indicate that boys are more variable than girls in terms of scores of specific cognitive tests (Dykiert, Gale, & Deary, 2009; Feingold, 1992; Hedges & Nowell, 1995) and general intelligence tests (Jensen, 1971). The validity of "greater male variability" hypothesis (Ellis, 1894) is still on the agenda of researchers. Ignoring the gender variability in efforts to explain success difference among different educational systems may lead to unrealistic interpretations. Therefore, students' success should be modelled with an appropriate approach that takes this variability into account.

Regression models which have been widely used in educational research intend to discover the direct relationship between success and a set of factors. These models have some strict assumptions such as homogeneity of variances, which, therefore, restrict the practicability in real-life problems. When the assumption of variance homogeneity among groups (e.g. girls and boys) is violated, the heteroscedasticity problem arises. In the existence of heterogeneity, coefficient estimates obtained from the standard linear regression remain unbiased, whereas in the logit regression, estimated coefficients are inefficient (no longer have the properties of unbiased and having minimum variance). Allison (1999) proposed a procedure to overcome the heterogeneity problem among groups, and later by considering the weaknesses of Allison's model, Williams (2009) introduced the heterogeneous choice models (HCM) which are a generalized version of logit and probit models (Rohwer, 2015). HCM allow testing and controlling the heterogeneity caused by gender difference, and therefore estimated coefficients which show the effects of factors on success can be comparable across countries. Besides, factors which have differential effects on success between girls and boys can be determined by adding interaction terms to HCM.

https://doi.org/10.33225/jbse/20.19.186

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Factors Affecting Science Success

Based on students' self-reported questionnaires prepared by International Association for the Evaluation of Educational Achievement (IEA), TIMSS data set provides a wide variety of information about different aspects of students' success such as student-related (e.g. gender and attitude towards science), family-related (e.g. parents educational level and socio-economic status) and school-related (e.g. class size and school location) factors (Hammouri, 2004; Shukakidze, 2013). According to the purpose, several research in recent years have been searched for the models that provide a better understanding of underlying factors on students' success for different content areas. While doing this, a different set of factors regarded as potentially effective on success has been considered. As was pointed out by Topçu, Erbilgin, and Arıkan (2016), finding the significant factors associated with students' success offers significant insights on low success and provides a guidance on which revisions are needed in the current curriculum to enhance success. To this end, a set of factors which can be seen as important causes of cross-national difference in students' science success was handled in this research.

Besides gender variable with its effect on success has been widely studied (Mullis & Stemler, 2002; Pavesic, 2008; Smith, Pasero, & McKenna, 2014; Thomson, 2008), student questionnaire includes a lot of information that can be used to explain the variation of success across students. Using Rasch partial credit model, different scales have been developed for the purpose of measuring a single underlying latent construct since the first cycle of TIMSS. Each scale covers different number of items that are grouped together (see, Martin et al., 2016b) and carries insights about different aspects of students' success such as their home learning environments, perceptions and attitudes toward science. Home Educational Resources (HER), Student Bullying (SB) and Students' Sense of School Belonging (SSB) are common developed scales for different content areas such as science and mathematics. On the other hand, Students Like Learning (SL), Students Confident (SC) and Students Value (SV) scales are created separately for each specific content. Except SSB, numerous studies have been conducted to investigate the importance of these scales on students' success (Kaya & Rice, 2010; Lai, Ye, & Chang, 2008; Smith, Pasero, & McKenna, 2014; Thomson, 2008; Tsai & Yang, 2015; Topçu, Erbilgin, & Arıkan, 2016). Since SSB scale is newly created, a limited number of studies dealing with its effect have been available (Hu, Leung, & Chen, 2018; Smith et al., 2019). Besides all these above-mentioned factors, the items namely "how far in education do you expect to go" (HF) is concern with the students' educational expectation and the item namely "weekly time spent on science homework" (SWKHW) have been regarded as significant factors related to students' success (Lee & Stankov, 2018; Sandoval-Hernández & Białowolski, 2016).

Research Focus

This research particularly focused on 8th grade students' science success and its determinants. The last-released data set was considered. The data set covered science results of all countries that were participated in TIMSS 2015. Thus, the interpretations were made for 39 countries simultaneously instead of a group of countries covering a specific geographical region or sharing similar educational practices. From this point of view, it can be said that a very comprehensive comparison of countries in modelling students' science success was provided with this research. Moreover, by considering the possible gender variability on success, the heterogeneity problem was tested and controlled via HCM. Different models were estimated according to the proposed modelling strategy for each country. Also, this research put forth the factors which significantly differed across gender groups in the effects of students' science success.

In this research, three research questions were being fully addressed for all 2015 TIMSS participating countries:

1. Among 39 participants, which country's success is affected by unobserved heterogeneity caused by gender differences?
2. What are the significant factors on science success? In which countries, the effect sizes of these factors on success take the highest and lowest values?
3. Which factor has significantly differential effect on students' science success across gender groups?

https://doi.org/10.33225/jbse/20.19.186

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Research Methodology

General Background

The data employed in this research were obtained from one of the largest international assessment studies—TIMSS, from the official website of TIMSS & PIRLS. This website provides researchers the ability to access restricted data. Also, using extended data is subject to permission by IEA. The data used in this research was directly downloaded from the IEA's international database, since it is public available (IEA, 2019). In 2015, the 6th cycle of the IEA studies was conducted, and the results were published in 2017 February. More than 580,000 students participated in TIMSS 2015 (Martin et al., 2016a). TIMSS 2015 assessed the 8th grade students' science achievement of 39 "participating" countries and 6 "benchmarking" entities. A student questionnaire includes many questions to collect different information such as demographic characteristics, attitudes toward learning mathematics/science, home environment, etc. Each student completed this student questionnaire (Hooper, 2016).

Sample

Since the focus group of this research was 8th graders, the sample of this research was limited with only 8th grade students' information and their science results. Besides, only "participating" countries were handled, and thus 39 countries were studied. Any student who did not answer the interested questions and scales used in this research was omitted from the sample. Thus, the research was performed with a total number of 229,493 students' information. Also, the data of some countries includes some unavailable information about scales used in this research. Details are explained at the end of Appendix A.

The raw data set includes five plausible values for each content area which are relied on Item Response Theory scaling to summarize students' science success (Foy, 2017). The first plausible value is the most common used among others, and therefore this research was based on the first standardized score. The standardized score has the mean value of 500 and the standard deviation value of 100. When the mean value is taken as center point, it is possible to classify students between those who have scored below 500 and those who haven't. Thus, a binary dependent variable which reflects a student's science success was created. When a student's science score was higher than the center point, the dependent variable was encoded as 1, and as 0 otherwise.

Gender information is a response of question "are you a girl or a boy?" and is measured as binary form (*girl*=1 or *boy*=2). To assess the effect the students' future educational goals on success, HF scale that is categorized under 3 levels (*post-graduate*=1, *university*=2, *less than university*=3) was included. HER is formed according to the students' degree of agreement to three items associated with the availability of home resources and measured in 3 levels of Likert rating scale (*many resources*=1, *some resources*=2, *few resources*=3). SB is based on the responses of 9 items concerning the students' experiences with bullying at school and the scoring is done on 3 point Likert rating scale (*almost never*=1, *about monthly*=2, *about weekly*=3). SSB has been started to compute in 2015 and according to the TIMSS developer team, it aims to reflect the students' feelings towards their school and connectedness with the school community (Martin et al., 2016b). SSB has three levels (*high sense of school belonging*=1, *sense of school belonging*=2, *little sense of school belonging*=3). SL is also based on 9 items in student questionnaire measures the affective (emotional) and behavioral reactions concerning interest in learning science (Kadijevich, 2008) by 3 levels of Likert rating scale (*very much like learning*=1, *like learning*=2, *do not like learning*=3). Another scale which measures students' attitudes toward science is SC and it is formed with 9 items. Students' self-appraisal of their ability to do well in science is measured by 3 levels (*very confident*=1, *confident*=2, *not confident*=3) (Smith et al., 2019). The last scale, SV, is created based on the responses of 9 items and mainly measures the students' thinking on the importance and usefulness of science (*strongly value*=1, *value*=2, *do not value*=3). Lastly, SKHW scale, created for science, was considered and the 3 levels of this scale were defined as: *3 hours or more*=1, *more than 45 minutes but less than 3 hours*=2, *45 minutes or less*=3.

Instrument and Procedures

HCM test and control of the unobserved heterogeneity by simultaneously estimating two equations, a choice and a variance, where the choice equation (Eq.1) reflects the effects of factors on choice (i.e. dependent variable)

https://doi.org/10.33225/jbse/20.19.186

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

and the variance equation (Eq.2) reflects the residual variability across groups (Keele & Park, 2006; Williams, 2009; Williams, 2010).

$$y_i^* = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i \qquad (1)$$

$$\sigma_i^2 = \exp(\mathbf{z}_i\boldsymbol{\gamma})^2 \qquad (2)$$

In the choice equation, $\mathbf{x}_i$ is a vector of factor values for the i^{th} observation and $\boldsymbol{\beta}$ is a vector of associated unknown regression parameters being estimated. $y_i^* y_i^*$ denotes the latent version of observed value of y_i which equals to 1 if $y_i^* > \tau$ (τ is the threshold parameter), and 0 otherwise. Standard errors of estimates are obtained assuming the residual term $\varepsilon_i \varepsilon_i$ logistic or normal distributed. In the variance equation, $\mathbf{z}_i$ is a vector of grouping factors for heteroskedastic groups in the underlying latent variable. $\boldsymbol{\gamma}$ $(= \ln \sigma)$ is a vector of unknown heteroskedasticity parameters and shows the degree of effect of $\mathbf{z}_i$ on variance.

Data Analysis

For the purpose of this research, models which test and control the unobserved heterogeneity caused by gender differences were estimated. The steps of modelling strategy for any country were introduced below:

1. The raw data set was cleaned from missing and/or not-available observations and the dependent variable denoting the corresponding success status of each student was created. Since the center point is equal to 500, the threshold parameter was taken as the center point.
2. Standard logit model with a set of factors (i.e. gender, HER, SB, SSB, SV, SL, SC, HF, SWKHW) was estimated.
3. 3. One additional interaction term (gender×HER) was added and the logit model with interaction term was estimated. Likelihood ratio (LR) test was then utilized to investigate the significance of interaction effect. The significance level was set at 5%. This step was repeated considering the interaction of each factor with gender (i.e. gender×SB, gender×SSB, gender×SV, gender×SL, gender×SC, gender×HF and gender×SWKHW). Thus, significant interactions were determined.
4. In order to control gender variation in unobserved heterogeneity, it was estimated two HCM models sequentially, starting with HCM without interactions, and then by including interaction terms specified in Step 3 (HCM with interaction effects). Gender was used as a grouping variable in both choice and variance equations to estimate two different HCM—without and with interaction effects. While HCM without interaction effects allowed investigating factors that affect students' success, HCM with interaction effects also provided additional search related with factors having gender-specific effects.
5. According to the LR test, the model which provided a much better fit to the data than does other was preferred and was reported.

Research Results

Estimates of heteroscedasticity parameters are given in Table 1. A negatively signed and significant estimation of $\ln \sigma$ in the variance equation implied that, after controlling for other factors, boys were characterized by lower variability in their science success. According to Table 1, standard deviation of the residuals in science success was exp(-.29)=.74 times lower for Bahraini boys than girls. Similar findings were evident for Kuwait, Norway, Oman and Turkey, which indicated that standard deviation of the residuals in science success was .73, .84, .80 and .87 times lower for boys than girls, respectively. Among these countries, the success variability was the largest across Kuwaiti girls and boys, while it was smallest across Turkish girls and boys. Contrary to this, HCM models estimated for Botswana, Japan, Lithuania, Singapore and United States indicated significant positive estimates of $\ln \sigma$ and it could be said that standard deviations of the residuals were 1.20, 1.23, 1.19, 1.17 and 1.20 times larger for boys' science success than girls, respectively. Among these countries, while the success variability was the largest between girls and boys in Japan, the variability was the smallest across Singaporean girls and boys.

As it was shown, residual variances in science success differed among gender groups for 10 countries. Therefore, interpretations about standard logit estimates would be biased for these countries when the heterogeneity problem is not identified and controlled. By controlling the unobserved heterogeneity caused by gender difference, Appendix A presents the HCM estimation results built for each country in modelling students' science success.

https://doi.org/10.33225/jbse/20.19.186

Table 1
Gender variability on success

Country	Country Code	*N*	Heterogeneity	
			ln(σ)	*t*-stat.
Australia	AUS	8876	-.02	-.40
Bahrain	BHR	4049	-.29***	-3.31
Botswana	BWA	5382	.19*	2.49
Canada	CAN	8036	.01	.18
Chile	CHL	4503	.08	1.10
Chinese Taipei	TWN	5634	.13	1.95
Egypt	EGY	6839	-.01	-.14
England	ENG	4391	-.10	-1.25
Georgia	GEO	3675	-.01	-.20
Hong Kong SAR	HKG	4031	.16	1.78
Hungary	HUN	4810	-.02	-.29
Iran, Islamic Rep. of	IRN	5786	.07	1.04
Ireland	IRL	4317	-.01	-.07
Israel	ISR	3945	-.09	-1.29
Italy	ITA	4138	-.11	-1.22
Japan	JPN	4299	.21*	2.10
Jordan	JOR	7067	-.07	-1.10
Kazakhstan	KAZ	4715	.01	.03
Korea, Rep. of	KOR	5244	.04	.59
Kuwait	KWT	3844	-.31*	-2.54
Lebanon	LBN	3257	.03	.27
Lithuania	LTU	4206	.18*	2.28
Malaysia	MYS	9120	-.09	-1.77
Malta	MLT	3403	.01	.09
Morocco	MAR	11340	-.06	-.68
New Zealand	NZL	7414	.01	.2
Norway	NOR	4159	-.17*	-2.29
Oman	OMN	7699	-.22**	-2.61
Qatar	QAT	4986	-.15	-1.93
Russian Federation	RUS	4704	-.01	-.04
Saudi Arabia	SAU	3294	-.08	-.60
Singapore	SGP	6010	.16*	2.03
Slovenia	SVN	4051	-.01	-.03
South Africa	ZAF	10627	.11	1.57
Sweden	SWE	3787	-.15	-1.48
Thailand	THA	6202	-.02	-.36
Turkey	TUR	5718	-.13*	-2.03
United Arab Emirates	ARE	16278	-.07	-1.77
United States	USA	9657	.19*	2.46

*Note: Bold text refers to significant estimates at * p< .05, ** p< .01 and *** p< .001.*

https://doi.org/10.33225/jbse/20.19.186

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Main Effects

The results of models preferred according to the modelling strategy described above are given with Appendix A. The first panel of the table shows the "Main Effects" and can be used in interpreting the effects of factors on science success. According to the table, it could be seen that students' science success was affected by gender in 9 countries. While negative estimates for Bahrain, Hong Kong SAR, Jordan, Korea and Oman suggested that boys had lower success than girls, positive estimates for Georgia, Italy, Lithuania and United States indicated that the odds of being successful for boys were higher than girls. When estimated odds ratios (ORs) were used to assess the strength of the relation of gender with science success, it was observed that the effect of gender on success was the strongest in Korea $\left(\hat{\beta} = -1.93, \mathrm{OR} = .15\right)$, while the weakest in Jordan $\left(\hat{\beta} = -.21, \mathrm{OR} = .81\right)$ among countries where success was in favor of girls. In other words, the success gap between gender groups was widest in Korea, while it was narrowest in Jordan. On the other hand, among countries where the science success was in favor of boys, gender affected the science success with the strongest magnitude in United States $\left(\hat{\beta} = .53, \mathrm{OR} = 1.70\right)$ and with the weakest magnitude in Italy $\left(\hat{\beta} = .30, \mathrm{OR} = 1.35\right)$.

Future educational goals seemed very effective on students' science success in all countries. Coefficient estimates of the third level of HF (i.e. *less than university*) were negative and statistically significant in all countries. This indicated that students who had educational expectations as *less than university* tended to be less successful in science as compared to students who had *post-graduate* plans. Estimated ORs ranged from .04 (Chinese Taipei) and .75 (Sweden) and, hence it could be said that the negative effect of having lower educational expectation on success was the strongest for Chinese Taipei, while was the weakest for Sweden.

HER was another effective factor on students' science success that was common for all participating countries. According to the estimates of the third level of HER (i.e. *few resources*), ORs ranged from .04 (South Africa) to .66 (Kazakhstan). This showed that negative effect of having *few resources* on science success was the strongest in South Africa, but weakest in Kazakhstan.

SSB was found to have significant effect on students' science success for 26 countries. The direction of the effect of this attitude on success varied from country to country. When the estimated coefficients of the third level of SSB (i.e. *little sense of school belonging*) were considered, lower sense of belonging was associated with lower success in students from Australia, Canada, Chile, England, Hong Kong SAR, Hungary, Ireland, Lebanon, Malta, New Zealand, Norway, Qatar, Sweden, United Arab Emirates and United States. Moreover, the difference in success between students with *little* and *high sense of belonging* was less pronounced among students in Hungary $\left(\hat{\beta} = -.31, \mathrm{OR} = .73\right)$, but more pronounced among students in United Arab Emirates $\left(\hat{\beta} = -.98, \mathrm{OR} = .38\right)$. On the other hand, estimates of the second level of SSB (i.e. *sense of school belonging*) were positive in Botswana, Egypt, Israel, Italy, Jordan, Malaysia, Oman, Singapore, South Africa, Thailand and Turkey. This suggested that compared to students with *high sense of school belonging*, students with *sense of school belonging* were more likely to succeed for these countries.

Except Bahrain, Chile, England, Hungary and Russian Federation, experiences of bullying were significantly linked with students' science success in all countries. According to the estimates of the third level of SB (i.e. *about weekly*), ORs lied in the range between .15 (Japan) and .81 (Morocco). Students who experienced bullying from their peers at school almost every week were less likely to succeed than those who *almost never* experienced bullying for these countries. In addition, the success gap between students with and without experiences of bullying was the widest in Japan, while the narrowest in Morocco. On the other hand, positive estimates were observed only for the second level of SB (i.e. *about monthly*) in Hong Kong SAR, Korea, Kuwait and Malta. Students from these countries who were bullied *about monthly* had higher scores than students who were *almost never* bullied.

SL was an effective factor on students' science success for 16 countries. Based on both the two levels of SL scale (i.e. *like learning* and *do not like learning*), results showed that the lower the tendency of this attitude, the lower the science success performance in Chile, Chinese Taipei, Egypt, Hong Kong SAR, Korea, Malaysia, Oman, Qatar, South Africa and Thailand. Moreover, the lowest ORs of success were observed in Korean students who *like learning* $\left(\hat{\beta} = -1.26, \mathrm{OR} = .28\right)$ and *do not like learning* $\left(\hat{\beta} = -1.52, \mathrm{OR} = .22\right)$ science among students from these 10 countries. This indicated that the decrease of this attitude toward science adversely affected mostly the Korean students' science success. On the other hand, in Australia, Botswana, Canada, Israel, Jordan and New Zealand, corresponding coefficients estimates were positive and statistically significant. Students from these countries who *like learning* or *do not like learning* science were more likely to be successful than their peers who *very much like learning*.

https://doi.org/10.33225/jbse/20.19.186

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Confidence in science was another common effective factor related with students' science success for all participants of TIMSS 2015. According to the third level of SC (i.e. *not confident*), coefficient estimates were all significantly negative, suggested that success decreased with decreased confidence. Moreover, the highest OR was found for Malaysian students which showed the effect of unconfident on success was the lowest in Malaysia $\left(\hat{\beta} = -.5, \mathrm{OR} = .61\right)$. On the other hand, this attitude toward science had the strongest negative effect on success for Japanese students $\left(\hat{\beta} = -3.22, \mathrm{OR} = .04\right)$. The success gap between students who were *very confident* in science and those who were *not confident* in science was the widest in Japan.

Based on the coefficient estimates of the SV scale, the odds of succeeding were lower in students who *valued* or *did not value* science compared to the students who *strongly valued* science for Australia, Canada, Kazakhstan, Korea, Lebanon, Malta, Morocco, Singapore, Slovenia, Sweden, Thailand, United Arab Emirates and United States. On the other hand, the situation was different for the students studying in Chile, Chinese Taipei, Israel, Norway, South Africa and Turkey. Higher success was observed in students who placed less value on science compared to those with a strong valuing of science.

The last investigated factor was the students' weekly time spent in doing their science homework. According to the results of SWKHW scale, negative and significant estimates were found only in Chinese Taipei $\left(\hat{\beta} = -.58, \mathrm{OR} = .56\right)$ and Singapore $\left(\hat{\beta} = -1.02, \mathrm{OR} = .36\right)$. This suggested that students in Chinese Taipei and Singapore who dedicated less time (*45 minutes or less*) for doing their science homework were 44% and 64% less likely to be successful from their peers who spent more time (*3 hours or more*), respectively. On the other hand, in Botswana, Chile, Egypt, England, Iran, Ireland, Israel, Italy, Japan, Jordan, Malaysia, Norway, Oman, Qatar, South Africa, Turkey and United Arab Emirates, students who spent less than 3 hours a week on their homework showed higher success compared to the students who spent *3 hours or more*.

Interaction Effects

Since HCM without interaction effects was the preferred model, no single factor had a sufficient magnitude in differentiating students' science success between girls and boys from England, Georgia, Ireland, Israel, Italy, Kazakhstan, Kuwait, Lebanon, Lithuania, Malta, Morocco, Norway, Singapore, Slovenia, South Africa, Sweden and Turkey. For this reason, cells of the second panel in Appendix A which shows the "Interaction Effects" are empty for these countries. However, HCM with interaction terms was the best fitted one in modelling students' science success for the remaining 22 countries. All possible interaction effects were estimated according to the modelling strategy (Step 3 of modelling strategy) but only factors that are significantly varied by gender are reported in Appendix A.

The interaction between gender and HF was statistically significant only in Bahrain. This result showed that, among students whose level of education expectations was *less than university*, boys were advantaged in the likelihood of being successful in science, compared to their peers with high future educational expectations for future. HER was another effective factor that made a difference on the success across gender groups for Bahrain, Hong Kong SAR, Qatar and Russian Federation. Except in Russian Federation, positive and statistically significant coefficient estimates of gender×HER indicated that boys with *some* or *few educational resources* had higher odds of success than girls with the same level of resources. In Russia, on the contrary, boys having *few educational resources* were disadvantaged in the likelihood of succeeding in science, compared to their peers having *many resources*. The effect of SSB varied by gender for only two countries – Canada and Jordan. Boys who had *sense of school belonging* tended to be less successful than girls with the same feelings in Jordan. However, the effect of SSB was higher, in favor of Canadian boys. Gender difference in the effect of bullying on science success was found in favor of boys among all countries with statistically significant interaction estimates. In Canada, Chinese Taipei, Egypt, Iran, Japan, Malaysia, New Zealand, Oman, Saudi Arabia, Thailand and United States, boys who experienced bullying at school were more likely to have greater odds of better score compared to girls with the same experiences. Significant interaction effect between SL and gender on science success was observed for Botswana, Japan, Korea, Malaysia, New Zealand and United Arab Emirates and all estimates took the negative values except for Korea. This means that the effect of SL on success varied in gender groups and the magnitude of the effect was lower among boys except Korea. In Korea, on the contrary, the success advantage was in favor of boys against girls. In Chile and United States, the odds of being successful while having no confidence in the ability of doing well in science were lower for boys against girls. However, for the same levels of confidence (i.e. *confident* and *not confident*), Bahraini boys were more successful than Bahraini girls. The interaction of gender×SV indicated that the success gap be-

https://doi.org/10.33225/jbse/20.19.186

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

tween *valuing* and *strongly valuing* in science was narrower in Canadian boys than Canadian girls. Contrary to this, gender difference was apparent in SV factor, in favor of boys from Australia, Hong Kong SAR and Hungary. Finally, no significant interaction effect between gender groups and SWKHW was found, which indicated that time spent on science homework is not a significant predictor that differentiates girls' and boys' science success.

Discussion

Students' success and its determinants have long been considered critical in assessment and management of educational politics and have attracted researchers that used both quantitative and qualitative approaches. TIMSS has provided comparable results on different content areas for many countries since 1995. Many studies have been done using TIMSS results to investigate factors that lay behind the students' success. Cross-country comparisons have also been made to examine similarities/differences of different educational systems' success determinants. In efforts of comparing success determinants across countries, unobserved heterogeneity caused by different dimensions, such as individual, school and time-based characteristics should be considered in making accurate interpretations. Gender heterogeneity in success is an important phenomenon and have been empirically proven by recent researches based on the overall scores of large-scale assessment studies that provide outcomes from samples of students in different nations (e.g. Baye & Monseur, 2016; Gray et al., 2019; Liu, Alvarado-Urbina, & Hannum, 2019; Reilly, Neumann, & Andrews, 2019). Therefore, investigating effective factors on students' success by considering the gender variability may produce accurate and efficient results for researchers aiming to compare the dynamics of different educational systems.

In recent years, important contributions have been made to the scientific literature by quantitative researchers to overcome biased coefficients caused by the unobserved heterogeneity across groups being compared and it is recommended to use these approaches instead of conventional methods (Mustillo, Lizardo, & McVeigh, 2018). However, while limited studies have been in an attempt to identify significant factors associated with students TIMSS success by considering different dimensions of heterogeneity (e.g. Sandoval-Hernández & Białowolski, 2016), there is no study available in modelling the success by considering the heterogeneity caused by gender differences. The motivation of this study was to fill this gap. First and foremost, a very comprehensive cross-country comparison in students' science success using the last released TIMSS data covering 39 countries was presented. While doing this, countries that have gender-related variability were identified, not only based on overall scores, but also considering other factors in science success. Furthermore, by adding interaction terms with gender to models built for countries, factors in the differentiation of science success across gender groups were determined.

One of the major findings of this research was that the residual variability in success differed by gender for 10 countries' science results. Greater male variability was observed in Botswana, Japan, Lithuania, Singapore and United States. On the contrary, girls were more variable in their science success in Bahrain, Kuwait, Norway, Oman and Turkey. Therefore, modelling efforts of students' science success should be carried out rigorously, especially for countries having gender variability.

On the other hand, second notable finding covered the effective factors on students' science success. When the results about gender difference on success were tackled, the success varied between girls and boys in 9 countries. Since one of the educational goals stated in "Education 2030 Program" by United Nations Educational, Scientific and Cultural Organization (UNESCO) is to recognize the gender equality in education, the efforts in eliminating gender-based discrimination plays a vital role in achieving the right to education for all (UNESCO, 2015). Therefore, especially for countries in which the success difference was evident among girls and boys, these findings are of key importance in monitoring and restructuring the secondary education policies and practices in terms of ensuring gender equity. Besides, as it was expected, future educational goals, home educational resources and confidence were common factors that affect the students' science success for all TIMSS participant countries. These findings are consistent with previous studies which have been pointed out that students with lower educational expectations, lower educational resources or lower confidence tend to be less successful than their peers with higher expectations, resources or confidence (e.g. Bouhlila, 2011; Filiz & Öz, 2019; Hammori, 2004; Liu & Meng, 2010; Ogura, 2006; Ramirez, 2006; Topçu, Erbilgin, & Arıkan, 2016). The education system, which focuses on the following issues, might make a significant progress in achieving success: (1) improving efforts to increase students' confidence in science; (2) improving efforts to increase students' access to educational resources; and (3) improving efforts to increase students' motivation about their expectations for the future. Apart from these factors, the current research also presented the effects of other success determinants for each country in detailed.

https://doi.org/10.33225/jbse/20.19.186

In assessing how the factors considered in this research differentiate the students' science success by gender, interaction terms were included to the models. Findings demonstrated that none of the factors had differential effect on students' success across girls and boys in 17 countries. The effects of factors on science success, on the other hand, differed by gender in 22 countries. One of the notable findings was about students' bullying experiences in 11 countries (Canada, Chinese Taipei, Egypt, Iran, Japan, Malaysia, New Zealand, Oman, Saudi Arabia, Thailand and United States). In these countries, the estimates of interactions between gender and SB scale were found to be statistically significant with positive signs. This indicated that being victims of bullying had differential effect on science success by gender for these countries and boys who experienced bullying at school were more likely to have greater odds of better score compared to girls with the same experiences. Besides, among all factors, it was found that only the students' time spent on doing their science homework had no significant interaction effect with gender. The effects of other factors on science success by gender were evident and varied from country to country. These findings also support some of the important studies searched for the success determinants that differentiate the girls' and boys' success using TIMSS results (Smith, Pasero, & McKenna, 2014; Thomson, 2008; Yoo, 2018).

Conclusions

This research provides remarkable results for 8th grade students' science success in 39 countries related with gender heterogeneity, effective factors and the importance of these factors in differentiating success among girls and boys. Since a dimension of heterogeneity related with gender differences was tested and controlled, comparisons of success determinants across countries can be more accurately done compared to traditional approaches that may produce biased estimates. Using the findings of this research, researchers can assess the status of the factors affecting science success for their own country. One can also compare the direction and magnitude of the effects of these factors on science success for a specific country with that of other countries having high success, and thus can identify the reasons behind low achievement. Identification of effective factors on students' success can be seen as one of the important steps in efforts of reducing the educational underperformance.

Since a broad comparison of the science results of TIMSS participating countries were presented, findings of this research highlight the actions that need to be taken into consideration by educators and educational policy makers. Firstly, the issue of gender inequality is seen as a global social justice problem. In the countries where gender heterogeneity was found to be statistically significant, actions to ensure gender equality should be one of the primary objectives in the efforts of improving educational policies. Secondly, since the models which explain the relationship between success and its determinants were built by considering the effect of heterogeneity, cross-country comparisons can be accurately made. Thus, the underlying reasons for underperformance of a country of interest can be explored by comparing it with the results of high performing countries. Lastly, determining the factors which cause the success gap between girls and boys can be used as a useful instrument for the efforts of closing this gap and, hence increasing the academic performance.

Limitations

Although this research provides extensive comparative results across 39 countries' science results, some limitations are present. This research is only based on 8th grade students' science success. Students' mathematics success can be handled in a similar manner for future research. When the information coming from 4th graders is used, comparisons between countries' elementary educational systems can also be made in addition to the secondary education. Besides, practical considerations limit the inclusion of many factors that can be seen as effective predictors of students' science success. There are also other factors that were not used in this research. Students' science success is also related with different dimensions such as parents' education level, time spent in extra lessons and students' engagement in science. Future research about modelling science success can be made examining these factors in order to obtain different insights.

References

Allison, P. (1999). Comparing logit and probit coefficients across groups. *Sociological Methods & Research, 28* (2), 186-208. https://doi.org/10.1177/0049124199028002003

Baye, A., & Monseur, C. (2016). Gender differences in variability and extreme scores in an international context. *Large-scale Assessments in Education*, *4* (1), 1-16. https://doi.org/10.1186/s40536-015-0015-x

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Bouhlila, D. S. (2011). The quality of secondary education in the Middle East and North Africa: What can we learn from TIMSS' results? *Compare, 41* (3), 327-352. https://doi.org/10.1080/03057925.2010.539887

Dykiert, D., Gale, C. R., & Deary, I. J. (2009). Are apparent sex differences in mean IQ scores created in part by sample restriction and increased male variance?. *Intelligence*, *37* (1), 42-47. https://doi.org/10.1016/j.intell.2008.06.002

Ellis, H. (1894). *Man and Woman: A study of human secondary sexual characters.* Walter Scott.

Feingold, A. (1992). Gender differences in mate selection preferences: A test of the parental investment model. *Psychological Bulletin*, *112* (1), 125. https://doi.org/10.1037/0033-2909.112.1.125

Filiz, E., & Öz, E. (2019). Finding the best algorithms and effective factors in classification of Turkish science student success. *Journal of Baltic Science Education, 18* (2), 239-253. https://doi.org/10.33225/jbse/19.18.239

Foy, P. (2017). *TIMSS 2015 user guide for the international database.* TIMSS & PIRLS International Study Center.

Gray, H., Lyth, A., McKenna, C., Stothard, S., Tymms, P., & Copping, L. (2019). Sex differences in variability across nations in reading, mathematics and science: a meta-analytic extension of Baye and Monseur (2016). *Large-scale Assessments in Education, 7* (2), 1-29. https://doi.org/10.1186/s40536-019-0070-9

Hammouri, H. (2004). Attitudinal and motivational variables related to mathematics achievement in Jordan: Findings from the Third International Mathematics and Science Study (TIMSS), *Educational Research*, *46* (3), 241-257. https://doi.org/10.1080/0013188042000277313

Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, *269* (5220), 41-45. https://doi.org/10.1126/science.7604277

Hooper, M. (2016). Developing the TIMSS 2015 context questionnaires. In M.O. Martin, I.V. Mullis, & M. Hopper (Eds.), *TIMSS & PIRLS 2015*, (pp. chapter 2, 1–8). TIMSS & PIRLS International Study Center.

Hu, X., Leung, F. K., & Chen, G. (2018). School, family, and student factors behind student attitudes towards science: The case of Hong Kong fourth-graders. *International Journal of Educational Research, 92*, 135-144. https://doi.org/10.1016/j.ijer.2018.09.014

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104* (1), 53-69.

International Association for the Evaluation of Educational Achievement (IEA) (2019). *TIMSS 2015 international database.* https://timssandpirls.bc.edu/timss2015/international-database/

Jensen, A. R. (1971). The race x sex x ability interaction. In R. Cancro (Ed.), *Intelligence: Genetic and Environmental Influences* (pp. 107-161). Grune and Stratton.

Kadijevich, D. J. (2008). TIMSS 2003: Relating dimensions of mathematics attitude to mathematics achievement. *Zbornik Institutaza Pedagoška Istraivanja/Journal of the Institute of Educational Research, 40* (2), 327-346. https://doi.org/10.2298/ZIPI0802327K

Karlson, K. B., Holm, A., & Breen, R. (2012). Comparing regression coefficients between same-sample nested models using logit and probit: A new method. *Sociological Methodology, 42* (1), 286–313. https://doi.org/10.1177/0081175012444861

Kaya, S., & Rice, D. C. (2010). Multilevel effects of student and classroom factors on elementary science achievement in five countries. *International Journal of Science Education, 32* (10), 1337–1363. https://doi.org/10.1080/09500690903049785

Keele, L., & Park, D. K. (2006). Difficult choices: An evaluation of heterogeneous choice models. In: *The 2004 meeting of the American Political Science Association*. (2-5 September 2004, Chicago, IL) (pp. 1-33). https://www3.nd.edu/~rwilliam/oglm/ljk-021706.pdf

Lai, S. H., Ye, R., & Chang, K. P. (2008). Bullying in middle schools: An Asian-pacific regional study. *Asia Pacific Education Review, 9* (4), 503-515. https://doi.org/10.1007/BF03025666

Lee, J., & Stankov, L. (2018). Non-cognitive predictors of academic achievement: Evidence from TIMSS and PISA. *Learning and Individual Differences, 65*, 50-64. https://doi.org/10.1016/j.lindif.2018.05.009

Liu, R., Alvarado-Urbina, A., & Hannum, E. (2019). Differences at the extremes? Gender, national contexts, and math performance in Latin America. *American Educational Research Journal*, 1-33. https://doi.org/10.3102/0002831219876236

Liu, S., & Meng, L. (2010). Re-examining factor structure of the attitudinal items from TIMSS 2003 in cross-cultural study of mathematics self-concept. *Educational Psychology*, *30* (6), 699-712. https://doi.org/10.1080/01443410.2010.501102

Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences*. Stanford University Press.

Martin, M. O., Mullis, I. V. S., Foy, P., & Hooper, M. (2016a). *TIMSS 2015 international results in science.* TIMSS & PIRLS International Study Center.

Martin, M. O., Mullis, I. V. S., Hooper, M., Yin, L., Foy, P., & Palazzo, L. (2016b). Creating and Interpreting the TIMSS 2015 Context Questionnaire Scales. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and Procedures in TIMSS 2015* (pp. 15.1-15.312). TIMSS & PIRLS International Study Center.

Martinez Abad, F., & Chaparro Caso López, A. A. (2017). Data-mining techniques in detecting factors linked to academic achievement. *School Effectiveness and School Improvement*, *28* (1), 39-55. https://doi.org/10.1080/09243453.2016.1235591

Mood, C. (2009). Logistic regression: Why we cannot do what we think we can do, and what we can do about it. *European Sociological Review, 26* (1), 67–82. https://doi.org/10.1093/esr/jcp006

Mullis I. V. S., & Stemler, S. E. (2002). Analyzing gender differences for high-achieving students on TIMSS. In D. F. Robitaille, & A. E. Beaton (Eds.), *Secondary analysis of the TIMSS data*. Springer.

Mullis, I. V. S., Martin, M. O., Fierros, E. G., Goldberg, A. L., & Stemler, S. E. (2000). *Gender differences in achievement: IEA's third international mathematics and science study (TIMSS).* TIMSS & PIRLS International Study Center.

Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics.* TIMSS & PIRLS International Study Center.

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). *TIMSS 2015 international results in mathematics.* TIMSS & PIRLS International Study Center.

https://doi.org/10.33225/jbse/20.19.186

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004a). *TIMSS 2003 international mathematics report: Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades.* TIMSS & PIRLS International Study Center.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004b). *TIMSS 2003 international science report: Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades.* TIMSS & PIRLS International Study Center.

Mustillo, S. A., Landerman, L. R., & Land, K. C. (2012). Modeling longitudinal count data: Testing for group differences in growth trajectories using average marginal effects. *Sociological Methods & Research*, *41*(3), 467-487. https://doi.org/10.1177/0049124112452397

Mustillo, S. A., Lizardo, O. A., & McVeigh, R. M. (2018). Editors' comment: A few guidelines for quantitative submissions. *American Sociological Review, 83* (6), 1281–1283. https://doi.org/10.1177/0003122418806282

Neuschmidt, O., Barth, J., & Hastedt, D. (2008). Trends in gender differences in mathematics and science (TIMSS 1995–2003). *Studies in Educational Evaluation, 34* (2), 56–72. https://doi.org/10.1016/j.stueduc.2008.04.002

Ogura, Y. (2006). Background to Japanese student achievement in science and mathematics. In S. J. Howie, & T. Plomp (Eds.), *Content of learning mathematics and science. Lessons learned from TIMSS* (pp. 313–331). Routledge.

Pavesic, B. J. (2008). Science achievement, gender differences, and experimental work in classes in Slovenia as evident in TIMSS studies. *Studies in Educational Evaluation, 34* (2), 94-104. https://doi.org/10.1016/j.stueduc.2008.04.005

Ramirez, M. (2006). Understand the low mathematics achievement of Chilean students: a cross-national analysis using TIMSS data. *International Journal of Educational Research, 45* (3), 102–116. https://doi.org/10.1016/j.ijer.2006.11.005

Reilly, D., Neumann, D. L., & Andrews, G. (2019). Investigating gender differences in mathematics and science: Results from the 2011 Trends in Mathematics and Science Survey. *Research in Science Education*, *49* (1), 25-50. https://doi.org/10.1007/s11165-017-9630-6

Rohwer, G. (2015). A note on the heterogeneous choice model. *Sociological Methods & Research, 44* (1), 145– 148. https://doi.org/10.1177/0049124114552750

Sandoval-Hernández, A., & Białowolski, P. (2016). Factors and conditions promoting academic resilience: a TIMSS-based analysis of five Asian education systems. *Asia Pacific Education Review, 17* (3), 511-520. https://doi.org/10.1007/s12564-016-9447-4

Schreiber, J. B. (2002). Scoring above the international average: A logistic regression model of the TIMSS advanced mathematics exam. *Multiple Linear Regression Viewpoints, 28* (1), 22–30.

Shukakidze, B. (2013). The impact of family, school, and student factors on student achievement in reading in developed (Estonia) and developing (Azerbaijan) countries. *International Education Studies*, *6* (7), 131-143. http://doi.org/10.5539/ies.v6n7p131

Smith, T. J., Pasero, S. L., & McKenna, C. M. (2014). Gender Effects on Student Attitude Toward Science. *Bulletin of Science, Technology & Society, 34* (1–2), 7–12. https://doi.org/10.1177/0270467614542806

Smith, T. J., Walker, D. A., Chen, H. T., & Hong, Z. R. (2019). Students' sense of school belonging and attitude towards science: A cross-cultural examination. *International Journal of Science and Mathematics Education*, 1-13. https://doi.org/10.1007/s10763-019-10002-7

Thomson, S. (2008). Examining the evidence from TIMSS: Gender differences in year 8 science achievement in Australia. *Studies in Educational Evaluation*, *34* (2), 73–81. https://doi.org/10.1016/j.stueduc.2008.04.003

Topçu, M., Erbilgin, E., & Arıkan, S. (2016). Factors predicting Turkish and Korean students' science and mathematics achievement in TIMSS 2011. *Eurasia Journal of Mathematics, Science and Technology Education*, *12* (7), 1711-1737. https://doi.org/10.12973/eurasia.2016.1530a

Tsai, L. T., & Yang, C. C. (2015). Hierarchical effects of school-, classroom-, and student-level factors on the science performance of eighth-grade Taiwanese students. *International Journal of Science Education, 37* (8), 1166-1181. https://doi.org/10.1080/09500693.2015.1022625

Tutz, G. (2018). Binary response models with underlying heterogeneity: Identification and interpretation of effects. *European Sociological Review*, *34* (2), 211-221. https://doi.org/10.1093/esr/jcy001

United Nations Educational, Scientific and Cultural Organization (UNESCO) (2015). *Education 2030: Incheon Declaration and Framework for Action for the Implementation of Sustainable Development Goal 4: Ensure inclusive and equitable quality education and promote lifelong learning.* http://uis.unesco.org/sites/default/files/documents/education-2030-incheon-framework-for-action-implementation-of-sdg4-2016-en_2.pdf

Williams, R. (2009). Using heterogeneous choice models to compare logit and probit coefficients across groups. *Sociological Methods & Research, 37* (4), 531–559. https://doi.org/10.1177/0049124109335735

Williams, R. (2010). Fitting heterogeneous choice models with oglm. *The Stata Journal, 10* (4), 540–567. https://doi.org/10.1177/1536867X1101000402

Wößmann, L. (2003). Schooling resources, educational institutions and student performance: The international evidence. *Oxford Bulletin of Economics and Statistics*, *65* (2), 117–170. https://doi.org/10.1111/1468-0084.00045

Wößmann, L. (2005a). Educational production in East Asia: The impact of family background and schooling policies on student performance. *German Economic Review, 6* (3), 331–353. https://doi.org/10.1111/j.1468-0475.2005.00136.x

Wößmann, L. (2005b). The effect heterogeneity of central examinations: Evidence from TIMSS, TIMSS-Repeat and PISA. *Education Economics*, *13* (2), 143-169. https://doi.org/10.1080/09645290500031165

Yoo, Y. S. (2018). Modelling of factors influencing gender difference in mathematics achievement using TIMSS 2011 data for Singaporean eighth grade students. *Asia Pacific Journal of Education*, *38* (1), 1-14. https://doi.org/10.1080/02188791.2017.1334626

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Appendix A. Results of modelling strategy

Country	AUS	BHR	BWA	CAN	CHL	TWN	EGY	ENG	GEO	HKG
Main Effects[A]										
Gender: Boy	.07	-1.75***	.25	.12	.39	.18	.04	-.15	.51***	-1.22**
HF: University	.01	-.01	-.40**	.05	-4.46***	-1.46***	.01	-.10	-.61***	.01
HF: Less than university	-1.09***	-.78***	-1.92***	-.92***	-1.77***	-3.13***	-1.25**	-.66***	-1.45***	-1.38***
HER: Some resources	-.99***	-.93***	-1.20***	-.89***	-1.49***	-1.16***	-.22	-1.46***	-.49***	-.73**
HER: Few resources	-2.56***	-1.31***	-2.30***	-1.78***	-2.72***	-2.03***	-.83***	-2.20***	-1.98***	-1.33***
SSB: Sense of school belonging	-.18**	.01	.40***	-.26**	-.24**	-.02	.33***	-.11	-.13	-.27**
SSB: Little sense of school belonging	-.58***	-.01	.69***	-.50***	-.65***	-.29	.27	-.50***	-.12	-.60***
SB: About monthly	-.12*	.01	-.21*	-.28***	-.03	-.12	-.60***	-.02	-.13	.42***
SB: About weekly	-.52***	-.02	-.56***	-.59***	-.47	-1.18*	-1.14***	-.24	-1.41***	.14
SL: Like learning	.19**	.01	.15	.24***	-.10	-.24	-.31***	.14		-.15
SL: Do not like learning	.14	-.01	.57*	.27**	-.32*	-.51*	-.21	.03		-.45**
SC: Confident	-.65***	-.72***	-1.06***	-.69***	-.58***	-.76*	-.75***	-.61***		-.57**
SC: Not confident	-1.36***	-1.41***	-1.56***	-1.38***	-.83***	-1.74***	-1.44***	-1.41***		-1.37***
SV: Value	-.27**	-.01	-.01	.16	.27**	.58**	.16	.09	-.09	-.02
SV: Do not value	-.46***	-.01	.45	-.23*	.56***	.54**	-.04	-.08	-.23	-.05
WKHW: 45 minutes - 3 hours	-.06	.01	.56**	.03	.33	.04	.44**	.67*		.25
WKHW: 45 minutes or less	-.08	.01	.78***	.01	.68**	-.58**	.51***	.16		.07
Interactions										
Boy × Less than university		.70*								
Boy × Some resources		.92***								
Boy × Few resources		1.30***								.82*
Boy × Sense of school belonging				.27*						
Boy × About monthly				.29*		.63*	.48**			
Boy × Like learning			-.51*							
Boy × Do not like learning			-1.29**							
Boy × Confident		.68**								
Boy × Not confident		1.32**			-.76**					
Boy × Value	.35**			-.30*						.56*
Boy × Do not value										.63**
Constant	-3.16***	-1.79***	-.72*	-2.78***	-1.31***	-6.10***	.86***	-3.27***	-.08	-3.55***
Log-likelihood	-4655.3	-2344.5	-2311.9	-4202.8	-2523.6	-2032.8	-2616.6	-2192.0	-1933.7	-1868.1

Country	HUN	IRN	IRL	ISR	ITA	JPN	JOR	KAZ	KOR	KWT
Main Effects[A]										
Gender: Boy	.23	.01	-.08	.10	.30***	.55	-.21*	.019	-1.93**	.02
HF: University	-1.03***	-.15*	.01	-.39***	-.07	-.17	-.24***	.02	-.13	-.41***
HF: Less than university	-2.64***	-1.20***	-.46***	-1.45***	-.64***	-1.38**	-1.55***	-.51***	-.93***	-1.16***
HER: Some resources	-.92***	-1.11***	-1.36***	-1.08***	-.75***	-1.18***	-.47***	-.24*	-.56***	-.38*
HER: Few resources	-2.67***	-2.25***	-2.41***	-2.93***	-1.70***	-2.49***	-1.08***	-.42**	-1.63***	-.93***
SSB: Sense of school belonging	-.12	.06	-.05	.21**	.25***	-.02	.40***	-.13	.01	.06
SSB: Little sense of school belonging	-.31*	.01	-.41**	-.20	.09	-.25	.73***	.15	-.26	-.26
SB: About monthly	.12	-.28**	-.08		-.13	-.20	-.19**	-.09	.30**	.20*
SB: About weekly	-.27	-.49	-.46*		-.64**	-1.87**	-.57***	-.57*	.38	-.21
SL: Like learning		-.01	-.01	.34**	.16	.09	.03		-1.26*	-.01
SL: Do not like learning		.20	-.16	.45**	.09	-.25	.30*		-1.52*	.06
SC: Confident		-.71***	-1.16***	-.97***	-.68***	-2.04*	-.91***		-.53	-.64***
SC: Not confident		-1.25***	-2.11***	-1.80***	-1.28***	-3.22***	-1.56***		-1.82***	-1.15***
SV: Value	-.01	.13	.15	.32**	.14	.05	.11	-.26***	-.66*	.12

https://doi.org/10.33225/jbse/20.19.186

SV: Do not value	-.12	.19	.01	-.01	.16	-.36	.03	.03	-1.10***	-.01
WKHW: 45 minutes - 3 hours		.26*	.37*	.55**	.25	1.13*	.45**		-.02	.25
WKHW: 45 minutes or less		.17	.41**	.74***	.38**	1.34**	.58***		.26	-.05
Interactions										
Boy ×Little sense of school belonging							-.89**			
Boy × About monthly		.33*								
Boy × About weekly						1.97*				
Boy × Do not like learning						-1.44**			1.50*	
Boy × Value	.42*									
Boy × Do not value	.53*									
Constant	-3.34***	-1.29***	-3.42***	-1.62***	-1.19***	-5.76***	-.01	-1.33***	-5.61***	.10
Log-likelihood	-2266.6	-3202.9	-2086.0	-2101.9	-2532.2	-1595.2	-337.2	-2936.7	-2389.3	-1726.8

Country	LBN	LTU	MYS	MLT	MAR	NZL	NOR	OMN	QAT	RUS
Main Effects[A]										
Gender: Boy	.11	.35***	.01	.03	.27	.14	-.11	-.23***	-.43	.32
HF: University	-.22	-.35**	-.28***	.01	-.45***	.38***	-.17	-.02	-.03	-.28*
HF: Less than university	-1.43***	-2.06***	-1.35***	-1.04***	-1.11***	-.70***	-.59***	-.72***	-.63***	-1.50***
HER: Some resources	-.27	-1.21***	-.97***	-.87***	-.95***	-1.47***	-.85***	-.57***	-.95***	-.82***
HER: Few resources	-1.30***	-2.40***	-1.68***	-1.85***	-1.42***	-3.07***	-2.00***	-.78***	-2.21***	-.71*
SSB: Sense of school belonging	.05	-.04	.27***	-.168	.09	-.07	-.06	.10*	-.02	.01
SSB: Little sense of school belonging	-.48*	-.07	-.06	-.43***	.11	-.51***	-.34*	.07	-.37***	-.22
SB: About monthly	-.38***	-.06	-.48***	.22*	-.17**	-.22**	-.06	-.08	.01	-.05
SB: About weekly	-1.31***	-.48*	-1.26***	-.38*	-.21*	-.47***	-.38*	-.37***	-.62***	-.11
SL: Like learning			-.28***			.30**	.03	-.16**	-.22	
SL: Do not like learning			-.60***			.23	-.02	-.25*	-.54***	
SC: Confident			-.78***			-.62***	-.94***	-.36***	-.64***	
SC: Not confident			-.50***			-1.29***	-1.79***	-.65***	-1.03***	
SV: Value	-.31**	.10	.20**	-.79***	-.27***	.10	.25**	.04	.05	.07
SV: Do not value	-.50*	-.02	-.82***	-1.40***	-.18	-.06	.28*	-.02	-.09	.15
WKHW: 45 minutes - 3 hours			.33***			.03	.60**	.36*	.61***	
WKHW: 45 minutes or less			.11			-.26	.62***	.49***	.25	
Interactions										
Boy × Some resources									.46*	
Boy × Few resources									.91*	-1.06*
Boy × About monthly			.39***			.32***		.29**		
Boy × About weekly			.52**							
Boy × Like learning						-.42**				
Boy × Do not like learning			-.65*							
Constant	.47**	-2.07***	-2.70***	-2.19***	.41**	-3.16***	-1.52***	-.34*	-1.11***	-2.51***
Log-likelihood	-1528.6	-2487.9	-5036.4	-1949.4	-3776.9	-4073.9	-2358.3	-4657.3	-2856.9	-2439.8

Country	SAU	SGP	SVN	ZAF	SWE	THA	TUR	ARE	USA
Main Effects[A]									
Gender: Boy	-.36	.19	.09	-.04	-.14	-.05	-.05	.10	.53**
HF: University	.08	-.18	-.67**	.19	.11	-.38***	-.37***	.01	-.05
HF: Less than university	-1.05***	-1.28***	-1.69***	-1.51***	-.29**	-1.78***	-1.60***	-.84***	-.85***
HER: Some resources	-.75***	-1.94***	-.84***	-1.94***	-1.35***	-1.52***	-1.33***	-.75***	-1.30***
HER: Few resources	-1.43***	-2.95***	-2.20***	-3.15***	-2.73***	-2.24***	-2.07***	-1.64***	-2.21***
SSB: Sense of school belonging	.13	.28**	.21	.29***	-.25**	.32***	.31***	-.46***	-.20***
SSB: Little sense of school belonging	.05	-.02	-.08	.79***	-.71***	.22	.80***	-.98***	-.48***

https://doi.org/10.33225/jbse/20.19.186

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

SB: About monthly	-.44**	-.25**	-.08	-.53***	.11	-.06	-.16*	-.09*	-.11
SB: About weekly	-1.16*	-1.25***	-.71***	-1.21***	-.69***	-.42**	-.88***	-.67***	-.55***
SL: Like learning	.10	-.01		-.07		-.18*	.05	.01	-.01
SL: Do not like learning	-.05	-.05		-.38*		-.18	-.15	.01	.18
SC: Confident	-.72***	-.14		-.36***		-.61***	-.96***	-.67***	-.59***
SC: Not confident	-1.47***	-.57***		-.70***		-1.02***	-1.54***	-1.15***	-1.20***
SV: Value	.14	.09	-.06	.16	-.01	-.23***	.14*	-.22***	.07
SV: Do not value	-.05	-.39*	-.61***	.73***	-.33**	-.30*	.47***	-.20*	-.17*
WKHW: 45 minutes - 3 hours	.73	-.01		.92***		.07	.39***	.42***	.20
WKHW: 45 minutes or less	.66	-1.02***		.71***		-.07	.34**	.18*	.03
Interactions									
Boy × About monthly	.71**					.36**			.37**
Boy × About weekly						.38*			.53**
Boy × Like learning								-.26**	
Boy × Not confident									-.71**
Constant	.58	-5.31***	-3.49***	.19	-2.26***	-2.55***	-2.24***	-1.72***	-2.72***
Log-likelihood	-1264.5	-2118.0	-1951.1	-2893.4	-2168.3	-3399.9	-3231.6	-9393.5	-5331.5

**p < .05; **p < .01; ***p< .001*

[A] Reference categories are the first level of factors.

Remarks: (1) Detailed results of the selection of significant factors (Step 3) are available upon request. (2) Students' information about SB scale is not available in science data set of Israel. Therefore, models for Israel are built without SB scale. (3) Students' information about SL, SC and SWKHW are not available in science data sets of Georgia, Hungary, Kazakhstan, Lebanon, Lithuania, Malta, Morocco, Russian Federation, Slovenia and Sweden. Therefore, models for mentioned 10 countries are built without these three factors.

Received: *January 14, 2020* Accepted: *April 03, 2020*

Cite as: Askin, O. E., & Oz, 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

Öyküm Esra Aşkın	PhD, Assistant Professor, Department of Statistics, Faculty of Arts and Sciences, Yildiz Technical University, Davutpasa Campus, Istanbul, Turkey. E-mail: oykumesra@gmail.com Website: https://avesis.yildiz.edu.tr/oeyigit/
Ersoy Öz *(Corresponding author)*	PhD, Associate Professor, Department of Statistics, Faculty of Arts and Sciences, Yildiz Technical University, Davutpasa Campus, Istanbul, 34220, Turkey. E-mail: ersoyoz@yildiz.edu.tr Website: http://avesis.yildiz.edu.tr/ersoyoz/