



## Türkiye'nin PISA 2015 Fen Performansının ve İlişkili Değişkenlerin Hiyerarşik Doğrusal Modelleme İle İncelenmesi

Mustafa YILDIZ <sup>1</sup>, Eda ERDAŞ KARTAL <sup>2</sup>, Günkut MESCİ <sup>3</sup>

<sup>1</sup> Amasya Üniversitesi, Eğitim Fakültesi, mustfa.yildiz@gmail.com,  
<http://orcid.org/0000-0002-3139-2698>

<sup>2</sup> Kastamonu Üniversitesi, Eğitim Fakültesi, erdaseda@gmail.com,  
<http://orcid.org/0000-0002-1568-827X>

<sup>3</sup>Giresun Üniversitesi, Eğitim Fakültesi, gunkutmesci@gmail.com  
<http://orcid.org/0000-0003-0319-5993>

Gönderme Tarihi: 25.12.2019

Kabul Tarihi: 21.03.2020

Doi: 10.17522/balikesirnef.663737

*Özet* – Bu çalışmanın amacı, PISA 2015'in Türkiye'deki 9. ve 10. sınıf öğrencilerinin fen performansına ilişkin değişkenleri hiyerarşik doğrusal modelleme (HLM) yaklaşımı kullanarak araştırmaktır. Bulgular, öğrencilerin fen performansı ile demografik özellikleri arasındaki ilişkinin, fen performansının duyuşsal alan ve öğrenme ortamı ile arasındaki ilişkiden daha güçlü olduğunu ortaya koymuştur. Bu bulgulara dayanarak, öğrencilerin fen performansını geliştirmenin, performansı etkileyen değişkenleri dikkate alarak, özellikle de demografik değişkenler ve kapsayıcı okul sistemlerinin inşası bağlamında eşit fırsatlar sağlayabilen gelişmelerle mümkün olabileceği düşünülmektedir.

*Anahtar kelimeler:* fen performansı, hlm, pisa

Sorumlu yazar: Eda ERDAŞ KARTAL, Kastamonu Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Merkez / Kastamonu

### Geniş Özet

#### Giriş

Öğrencilerin bilimsel okuryazarlık seviyelerinin belirlenmesi, müfredattaki son değişikliklerin kapsamını ve Türkiye'deki eğitim reformlarının amaçlarını belirlemek açısından önemlidir. PISA (Uluslararası Öğrenci Değerlendirme Programı) eğitim ve öğretimi

izlemek için periyodik olarak veri toplayan uluslararası bir değerlendirme programıdır. Türkiye'de öğrencilerin ortalama bilim okuryazarlığı, Ekonomik İş birliği ve Kalkınma Örgütü (OECD) ülkeleri ortalamasının altındadır (MEB, 2003, 2006, 2009, 2012, 2018). Türkiye'nin fen müfredatı uluslararası arenadaki gelişmelere paralel olarak periyodik olarak güncellenmekte ve sınıf öğretmenlerinin uygulamaları çeşitli mesleki gelişim programları ile desteklenmektedir (MEB, 2005, 2013, 2018). Bununla birlikte, PISA testlerinin sonuçları, ağırlıklı olarak fen okuryazarlığı olan PISA sınavlarında (PISA 2006, PISA 2015), program reformlarının ve mesleki gelişim uygulamalarının Türkiye'deki öğrencilerin fen puanlarını artırmadığını göstermiştir (Tablo 1). Türkiye'nin fen performansını arttırmak için, yapılan uluslararası sınavlarda öğrencilerin fen performansını hangi değişkenlerin etkilediğinin belirlenmesi önem arz etmektedir. Bu çalışmanın amacı, öğrencilerin fen performansı ile öğrenme ortamı, duygusal özellikler ve demografik özelliklerle ilgili bir dizi değişken arasındaki ilişkiyi hiyerarşik doğrusal modelleme (HLM) yaklaşımı kullanarak incelemektir.

## Yöntem

Bu çalışmanın örneklemini PISA 2015 çalışmasının Türkiye kısmı oluşturmaktadır. 2015 yılında ülke genelinde 159 farklı liseye devam eden 5581 adet 9 ve 10. sınıf öğrencisi PISA sınavına katılım sağlamıştır. Bağımlı değişken, SCIENCE, öğrencilerin fen içerik bilgi düzeylerini belirlemek üzere tasarlanmış sürekli bir değişkendir. Değerlendirilen bağımsız değişkenler; duygusal özellikler, öğrenme ortamı ve demografik özellikler olarak üç ana kategoride gruplandırılmıştır (Tablo 2). Ölçeklerin güvenilirlik düzeyleri PISA raporundan elde edilmiştir (OECD, 2015a). Bağımlı ve bağımsız değişkenler arasındaki ilişki iki düzeyli HLM modeli kullanılarak incelenmiştir. Alt düzey öğrenci düzeyi, üst düzey ise okul düzeyi olarak belirlenmiştir. Sonraki aşamalarda biri boş olmak üzere toplam 5 model oluşturulmuş ve bu modellerin açıkladıkları varyans miktarları incelenerek, hangi değişken grubunun öğrenci fen performansını ne düzeyde etkilediği araştırılmıştır. Model 0, herhangi bir bağımsız değişken içermeyen boş modeldir. Bu model toplam varyansın ne kadarının grup üyeliğinden (aynı sınıfta öğrenci olmak) kaynaklandığını belirlemek için kullanılmıştır. Model 1, duygusal özellikleri içeren bağımsız değişkenlere sahiptir. Model 1'i kullanmanın amacı, duygusal özelliklerin fen başarısıyla ne kadar ilişkili olduğunu görmektir. Model 2, öğrenme ortamı ile ilgili özellikleri ölçen bağımsız değişkenleri içermektedir. Model 3, öğrenci ve okulların demografik özelliklerini içeren değişkenleri içermektedir. Model 4, önceki üç modelde yer alan tüm bağımsız değişkenleri içeren modeldir.

## **Bulgular**

Bulgular, öğrencilerin fen performansı ile demografik özellikleri arasındaki ilişkinin, fen performansı ile duyuşsal alan ve öğrenme ortamı arasındaki ilişkiden daha güçlü olduğunu ortaya koymuştur. Fen performansını açıklama noktasında değişkenlerin etki büyüklükleri en büyüğünden en küçüğüne sırayla şu şekildedir; sosyo-ekonomik düzey, okul türü, sınıf düzeyi, cinsiyet, sorgulama temelli öğretim etkinliklerinin kullanılması, epistemolojik inançlar, bilimden zevk alma, öz yeterlik, öğretimin adaptasyonu, fen konularına ilgi, öğretmen merkezli öğretim, öğretmen destek düzeyi, sınıftaki disiplin ortamı, motivasyon ve araçsal motivasyondur. Öğrencilerin demografik özelliklerinden sosyoekonomik durum, cinsiyet ve sınıf düzeyi ile bağımlı değişken olan fen performansı arasındaki ilişkilerin istatistiksel olarak anlamlı olduğu ortaya çıkmıştır. Ayrıca, öğrencilerin duyuşsal özelliklerinden epistemolojik inançlar, fenden zevk alma, öz yeterlik ve fen konularına ilgi gösterme değişkenleri ile fen başarısı arasındaki ilişkinin de anlamlı olduğu gözlemlenmiştir. Buna karşın, öğretim ortamının özellikleri ile ilgili değişkenlerden yalnızca fen etkinliklerine katılma düzeyi ile fen performansı arasındaki ilişkinin istatistiksel olarak anlamlı olmadığı fark edilmiştir. Bu kategorideki diğer değişkenlerle fen performansı arasındaki ilişkinin anlamlı olduğu tespit edilmiş, ancak bu değişkenler sayesinde açıklanabilen ilave varyans (boş model baz alındığında) 0.1 düzeyinin altındadır.

## **Sonuç ve Tartışma**

Bulgular öğrencilerin fen performanslarını oluşturan varyansın büyük bir kısmının Model 3 (öğrencilerin demografik özellikleri) ile açıklanabileceğini göstermiştir. Bu modeldeki iki değişkenin (TRATIO ve SCTYP) dışındaki değişkenlerin (SES, GENDER ve GRADE) öğrencilerin fen performansı ile istatistiksel olarak anlamlı bir ilişki içinde oldukları tespit edilmiştir. Sonuç olarak, Türkiye'deki fen performansının öncelikle demografik özellikler tarafından belirlendiği ileri sürülebilir. Buna ilave olarak, öğrencilerin fen dersine yönelik tutum ve inançlarının fen performanslarına olan etkisi öğrenme ortamının özelliklerine nispeten daha fazladır. Model 3'te bulunan değişkenlerden biri olan SES'in (sosyo ekonomik statü) etki büyüklüğü bariz bir şekilde diğer değişkenlerin önündedir. Bu bulgu literatürdeki diğer çalışmalarla tutarlılık göstermektedir (Alivernini & Manganelli, 2015; Sun et al. 2012). SES'e yakın etki büyüklüğüne sahip bir diğer değişken okul türü (SCTYP) değişkenidir. Bu değişkenin etki büyüklüğü, özel okullarda okuyan öğrenciler ile devlet okullarındaki öğrenciler arasındaki farkı yansıtmaktadır. Parametrenin istatistiksel olarak anlamlı olmamasının nedeni örnekleme yer alan okulların yalnızca % 4'ünün özel

okul olmasından kaynaklanmış olabilir. Bu bulgu Avustralya örneği ile tutarlı olsa da (Perry & McConney, 2004), Singapur örneği ile tutarlı değildir (Areepattamannil et al., 2015). Perry & McConney (2004) çalışmalarında Avustralya'daki yüksek sosyoekonomik düzeydeki öğrencilerin çoğunun özel okullarda öğrenim gördüğünü belirtmiştir. Okul türünün Türkiye'deki öğrencilerin fen performansına etkisi de okul türleri arasındaki imkan ve kaynak farklılığından kaynaklanıyor olabilir. Çalışmamız bu öngörüğü destekler niteliktedir. Buna ek olarak, OECD 2004 raporu, okul türleri arasında hiçbir fark bulunmayan daha kapsayıcı okul sistemlerinin (ör. Kanada ve Finlandiya) daha yüksek performans elde etmede etkili olacağını belirtmektedir. Etki büyüklüğüne göre SES ve SCTYP değişkenlerini sırasıyla GRADE (sınıf düzeyi) ve GENDER (cinsiyet) değişkenleri takip etmektedir. Araştırma bulgularına göre, erkek öğrencilerin fen performanslarının kız öğrencilerin fen performanslarından daha yüksektir. Literatür incelendiğinde, bu bulguların diğer çalışmaların bulgularıyla uyumlu olduğu bulunmuştur (Areepattamannil et al. 2015; Lam & Lau, 2014). Bu çalışmada cinsiyetle ilgili bulgular, Türkiye örnekleminin önceki PISA çalışmalarına göre farklılık göstermektedir. Fen performansı ve cinsiyet arasındaki ilişkiyi inceleyen Türkiye örneklemleri önceki çalışmalar, cinsiyet değişkeninin kız öğrenciler lehine fen performans farkına sahip olduğunu bildirmiştir (Akkuş, 2008; Gürsakal, 2012). Türkiye'de sosyoekonomik olarak dezavantajlı ailelerde kızların okula gitme oranlarının düşük olduğu bilinmektedir (Kocabaş Aladağ ve Yavuzalp, 2004). Bu çalışmada, dezavantajlı gruplardan PISA 2015 uygulamasına katılan kız öğrencilerin oranı, önceki PISA uygulamalarından daha düşüktür ve bu, erkek ve kız öğrenciler arasındaki farkın azalmasına neden olmuş olabilir.

## Öneriler

Bu çalışma, yeni hedefler ve reformlar hakkında Türkiye'ye rehberlik edecek önemli ipuçları ortaya koymaktadır. Bu bulgulara dayanarak, özellikle demografik değişkenler ve kapsayıcı okul sistemlerinin inşası bağlamında eşit fırsatlar sağlayabilen iyileştirmelerle, öğrencilerin fen performanslarını artırmanın, performansı etkileyen değişkenler dikkate alınarak mümkün olduğu düşünülmektedir. Politikacıların ve eğitim programcılarının bunu dikkate alması önerilmektedir.

# Investigation of Turkey's PISA 2015 Science Performance and Associated Variables Using Hierarchical Linear Modeling

**Mustafa YILDIZ<sup>1</sup>, Eda ERDAS KARTAL<sup>2</sup>, Günkut MESCİ<sup>3</sup>**

<sup>1</sup> Amasya University, Faculty of Education, mustfa.yildz@gmail.com,  
<http://orcid.org/0000-0002-3139-2698>

<sup>2</sup> Kastamonu University, Faculty of Education, erdaseda@gmail.com,  
<http://orcid.org/0000-0002-1568-827X>

<sup>3</sup>Giresun University, Faculty of Education, gunkutmesci@gmail.com  
<http://orcid.org/0000-0003-0319-5993>

Received: 25.12.2020

Accepted: 21.03.2020

Doi: 10.17522/balikesirnef.663737

---

*Abstract* – The purpose of this study is to investigate the variables related to science performance of 9<sup>th</sup> and 10<sup>th</sup> grade students in Turkey portion of PISA 2015, by using hierarchical linear modeling (HLM) approach. The findings revealed that the relationship between the science performance and demographic characteristics of the students is stronger than the relationship between science performance and the affective domain as well as the learning environment. Based on these findings, it is thought that improving students' science performance is possible by taking into account the variables that affect the success, especially with improvements that can provide equal opportunities in the context of demographic variables and the construction of inclusive school systems.

*Key words:* science performance, hlm, pisa

-----  
Corresponding author: Eda ERDAS KARTAL, Kastamonu University, Faculty of Education, City Center / Kasatmonu

## Introduction

Science literacy is defined as the understanding of science by individuals and using them in a scientific and technological discussion outside of school (Ryder, 2001). Developing scientifically literate individuals who can keep up with the rapid development and progress in science and technology in the globalized world is the main goal of the many science education programs (Ministry of National Education [MoNE, Turkey], 2018; Next Generation Science Standards [NGSS Lead States], 2013). The evaluation of the findings from various

assessment programs at international level is important in terms of determining the extent to which educational services reach their goals. PISA (Program for International Student Assessment) is one of the assessment programs, findings of which could be used for that purpose. Turkey has been participating in the PISA project since 2003. Average science literacy of students in Turkey was below the average of Organization for Economic Cooperation and Development (OECD) countries (MoNE, 2003, 2006, 2009, 2012). According to the last test results in 2018, although Turkey has increased the score in the area of scientific literacy is still below the average of OECD countries. In addition, as can be seen from the table below, the average of science literacy in exams in 2006 and 2015, which were predominantly examined for science literacy, has decreased compared to previous years.

**Table 1** Turkish Students' Science Literacy Average in PISA Tests

	PISA 2003	PISA 2006	PISA 2009	PISA 2012	PISA 2015	PISA 2018
Average	435	<b>424</b>	454	463	<b>425</b>	468

The bold averages show that the exam is predominantly science literacy.

The level of Turkish students' performance in PISA shows that the objectives of the latest educational reforms were not achieved (Acar & Ogretmen, 2012; Özdemir, 2010). Turkey's science curriculum is updated periodically in parallel with the developments in the international arena, and practices of teachers in the classroom are supported by a variety of professional development programs (MoNE, 2005, 2013, 2018). However, the results of PISA tests showed that program reforms and the implementations of professional developments failed to increase the level of students' science scores in Turkey in PISA exams which are predominantly science literacy. As a result, it becomes essential to revisit the variables on which new development programs delivered to see to what extent they are related to students' science scores.

When examining the literature on this subject, it was seen that there are limited analyses using directly the data from Turkey's PISA test. In a study was conducted by using the data from PISA, Özdemir (2017) examined 97 studies carried out using Turkish PISA data, and the results were very remarkable. In more than half of the studies analysis in Özdemir's (2017) study, it was reported that using PISA data performed no original analysis and the results of this studies had already published by OECD and MEB. In the majority of the articles in which the original analyzes were conducted, it was stated that the methodological requirements were not followed (e.g., sample weights, possible values, and software used for analysis) and thus

*Investigation of Turkey's PISA 2015 Science Performance and Associated Variables Using Hierarchical Linear Modeling*

the analyses were far from being reliable (Özdemir, 2017). The findings of Özdemir's (2017) study clearly demonstrate the necessity of the studies to be done using the original analysis with using Turkey's PISA data. This study seeks to examine a host of factors using HLM (hierarchical linear modeling) in order to give a more holistic picture to account for the science performances of Turkish students in PISA 2015.

*Literature Review*

In this section, the literature review related to a group of variables, which are measured in PISA evaluations and which can be related to students' performance, are presented under three main headings.

*Students' Affective Characteristics on Science Performance*

The findings of the studies conducted with PISA data show that variables about students' affective characteristics such as self-efficacy, enjoyment from the science, interest in science and motivation for success are related to student performance. For example, in a study using PISA 2006 Hong Kong data, it was reported that "self-efficacy" and "enjoyment from science", which are considered among attitudinal factor, play an important role for acquiring science objectives (Lam & Lau, 2014). In this context, another study conducted with PISA 2006 Hon Kong data revealed that students' science performance was significantly associated with students' motivation and self-efficacy (Sun, Bradley, & Akers, 2012). In a study in which the factors related to the success of East Asian countries were analyzed using the 2009 data, it was revealed that the effect of general interest in science learning on students' science performance was found to be positive relation at the country level, while the effect of students science topic interest was negative relation (Bybee & McCrae, 2011).

When the literature is examined, it can be seen that epistemological beliefs may be one of the affective characteristics, which may affect students' learning and performance (Muis, Bendix & Haerle, 2006). However, it is seen that the findings in the context of the relationship between scientific epistemological beliefs and student performance are inconsistent. For example, in a study that examined the relationship between students' epistemological beliefs and their performance in a comprehension test formed from PISA 2006 questions, it was revealed that there is no linear relationship between the students' scientific epistemological beliefs and their conceptual understanding (Sadıç & Çam, 2015). In another study, it was revealed that the epistemological beliefs about the development of

scientific knowledge directly affect the content knowledge (Mason, Boscolo, Tornatora, & Ronconi, 2013).

The instrumental motivation of the student (instrumental motivation to science learning) can also be considered as one of the affective characteristics associated with success (MoNE, 2016). Instrumental motivation is that students are willing to learn science and be willing to care for themselves and their future careers (Wigfield & Eccles, 2000). In a study in which PISA 2003 data was used to examine factors affecting student and school level that affect the mathematics performance of Korea, Japan and America, instrumental motivation was found to be one of the factors predicting student performance (Shin, Lee, & Kim, 2009). On the other hand, in a study using Turkey PISA 2015 data, it was established that there is no significant relationship between students' science performance and instrumental motivation (Yetiştir, Batı, Kahyaoğlu & Birel, 2018).

#### *Learning Environment Characteristics on Science Performance*

One of the characteristics of the learning environment is the disciplinary climate in class. The findings of studies investigating the effect of variables related to the learning environment on student performance are inconsistent. For example, in a study using the 2006 and 2009 data of 10 developed OECD countries (Korea, Japan, Australia, Germany, France, UK, USA, Italy, Spain), it was revealed that the disciplined classroom environment was positively related to student performance (Sousa, Park & Armor, 2012). However, a strict classroom discipline in the school does not guarantee high academic performance (Güzel & Berberoğlu, 2005). In contrast to countries such as Japan and the USA, which have high reliance on disciplinary action strategies, Finland has a lower disciplinary climate, but has higher rank in PISA assessments (Ning, Van Damme, Van Den Noortgate, Yang & Gielen, 2015).

One of the other characteristics of the learning environment is the way of teaching. Some of the variables related to how teaching is done such as inquiry-based or teacher-centered, students' science activities, the level of support of learning by the teacher, perceived feedback by the student, and the re-arrangement of the course according to needs. Inquiry-based teaching is defined as a more student-centered type of teaching where the teacher-led learning experiences are progressively reduced (Wise & Okey, 1983). Similarly, studies on the effect of preference of inquiry-based instruction on teaching on science performance are also inconsistent. Some studies reported a positive effect (Jiang & McComas, 2015; Minner,



*Investigation of Turkey's PISA 2015 Science Performance and Associated Variables Using Hierarchical Linear Modeling*

Levy & Century, 2010) while some studies reported negative effects (Cairns & Areepattamannil, 2017).

*Demographic Characteristics on Science Performance*

Studies show that the gender variable is one of the demographic variables associated with the success of the students (Areepattamannil, 2014; Lam & Lau, 2014). In a study conducted with PISA 2006 Hon Kong data, it was revealed that the students' science performance was significantly related to the gender of the students (in favor of male students) (Sun et al. 2012).

Research shows that socioeconomic status is one of the demographic variables related to student performance (Alivernini & Manganelli, 2015; Sun et al. 2012). In a study conducted using PISA 2003 Australian data, the impact of school socioeconomic level on students' performance was investigated, and it was revealed that increases in the average socioeconomic level of a school were related to consistent increases in students' academic performance and this relationship was similar for all students regardless of individual socioeconomic levels (Perry & McConney, 2004).

Type of school (in favor to public) and average number of students per teacher in a school are another demographic variables related to student performance. But the results of research on these variables are inconsistent. For example, in a study which PISA 2009 Singapore data were used, it was found that the type of school (public or private), number of students per teacher, quality of educational resources of the school, and student and teacher behaviors affecting school climate of school-level variables were not significantly related to students' science performance (Areepattamannil, Chiam, Lee & Hong, 2015). In contrast to this study, in a study evaluating the 2006 and 2009 data of 10 developed OECD countries, it is revealed that the type of school (in favor to public school) is positively related to student performance and the number of students per teacher is not related to student success (Sousa et al. 2012).

*Aim of the study*

The purpose of this study was to investigate the relationship between students' science performance, and a set of variables that are related to learning environment, affective characteristics, and demographics. The dataset use is the Turkey portion of PISA 2015. The aim of this study is to examine the 2015 PISA data, as this is the last exam that is focused on

science literacy field; therefore, variables related to science (eg affective domain) were measured in this exam.

## Method

### Sample

Turkey portion of the PISA 2015 was used as the sample of this study. There were 5581 students who attended 159 high schools across the country. Students were either 9<sup>th</sup> or 10<sup>th</sup> grade. The demographic characteristics of the students such as grade, gender, and school type are displayed in Table 2.

**Table 2** Characteristics of The Sample

School Type	Gender	9 <sup>th</sup> grade	10 <sup>th</sup> grade	Total
Vocational	Female	244 (4%)	843 (15%)	1087 (19%)
	Male	464 (8%)	914 (16%)	1378 (25%)
High school	Female	250 (4%)	1429 (26%)	1979 (35%)
	Male	315 (6%)	1122 (20%)	1437 (26%)
Total		1273 (23%)	4308 (77%)	5581 (100%)

### Variables

The dependent variable, SCIENCE, was measured as a continuous variable and it is the science content knowledge subscale of PISA assessment. Performance levels indicated by this variable were quantified using binary item response theory models (De Ayala, 2013). The independent variables considered were grouped into three main categories as affective characteristics, learning environment, and demographics (Table 2). The reliability of the scales was obtained from the PISA report (OECD, 2015a).

**Table 2** Independent Variables\*

Domain	Variable	Number of items	Scaling	Reliability
Affective characteristics	EPIST	6	Continuous IRT	0.92
	JOYSCI	5	Continuous IRT	0.94
	SCIEEFF	7	Continuous IRT	0.89
	INTBRSCI	4	Continuous IRT	0.85
	INSTSCIE	4	Continuous IRT	0.90
	MOTIVAT	5	Continuous IRT	0.84

*Investigation of Turkey's PISA 2015 Science Performance and Associated Variables Using Hierarchical Linear Modeling*

	DISCLISCI	5	Continuous IRT	0.89
	IBTEACH	8	Continuous IRT	0.89
	TEACHSUP	5	Continuous IRT	0.91
Learning environment	TDTEACH	4	Continuous IRT	0.80
	PERFEED	5	Continuous IRT	0.91
	ADINST	3	Continuous IRT	0.81
	SCIEACT	9	Continuous IRT	0.94
	GRADE	1	Binary	-
	GENDER	1	Binary	-
Demographics	SES	9	3 continuous subscales	0.68
	SCTYPE	1	Binary	-
	TRATIO	1	Continuous	-

\*EPIST: epistemological beliefs; JOYSCI: enjoyment of science; SCIEEFF: science self-efficacy; INTBRSCI: interest in broad science topics; MOTIVAT: student attitudes, preferences and self-related beliefs, achieving motivation; DISCLISCI: the disciplinary climate in a science class; IBTEACH: inquiry-based science teaching and learning practices; TEACHSUP: teacher support in a science class; TDTEACH: teacher directed science instruction; PERFEED: perceived feedback; ADINST: adoption of instruction; INSTSCIE: instrumental motivation; SCIEACT: students' science activities; SES: socio-economic status; GRADE: whether a student is a 9th grade or 10th grade; SCTYPE: whether a school is public or private; TRATIO: average number of students per teacher in a school; GENDER: male or female; HISEI: Highest parental occupational stat; PARED: Highest education of parents in years; HOMEPOS: home possessions including books in the home; IRT: Item response theory; Reliability: The reliability values are of Turkey sample.

The measurement scales within affective characteristics and learning environment categories had four-point Likert type items and were generated using partial credit item response theory models (Muraki, 1992).

*Data Analysis**Using Multilevel Models*

When the data has a nested structure such as students nested in schools, patients nested in clinics, kids nested in families, use of HLM rather than ordinary least squares methods is more appropriate (Finch, Bolin & Kelley, 2016). In PISA studies, the sampling method was a two-stage stratified sampling (OECD, 2015b) meaning that schools were randomly selected from the population of schools, then participants were selected from the school that were selected, which implies that the data has a nested structure. In this context, the use of a HLM would be more realistic since the assumption of *independence of observations* put forward by the traditional methods is violated.

A two-level HLM was used as the lower level was the student level, and the higher level was the school level. Then, the second step was to test a null model and four other models that had the random intercepts and fixed slopes (Snijder & Bosker, 1999) (Table 4).

Model 0 was the empty model that does not contain any independent variable, which is used to see how much of the total variance is accounted for by the group membership as well as the baseline model for comparison. Model 1 has the independent variables that were measures of the affective characteristics. The purpose of using Model 1 is to see how affective characteristics are related to science performance. Model 2 includes independent variables that measure characteristics related to the learning environment. Model 3 contains demographic variables of students and schools. Model 4 was the model containing all of the independent variables that were included in the previous three models. The interval or ratio level independent variables in each model were grand-mean centered so that the interpretation of the intercepts could be meaningful.

**Table 4** Models and The Independent Variables

Model*	Level-1 IVs	Level-2 IVs
Model 0	-	-
Model 1	EPIST, JOYSCI, SCIEEFF, INTBRSCI, MOTIVAT, INSTSCIE	-
Model 2	DISCLISCI, IBTEACH, TEACHSUP, TDTEACH, PEFEED, ADINST, SCIEACT	-
Model 3	GRADE, GENDER, SES	SCTYPE, TRATIO
Model 4	EPIST, JOYSCI, SCIEEFF, INTBRSCI, MOTIVAT, DISCLISCI, IBTEACH, TEACHSUP, TDTEACH, PEFEED, ADINST, INSTSCIE, SCIEACT, GRADE, GENDER, SES	SCTYPE, TRATIO

\*All of these models are intercepts-only models; IV: independent variable; Level-1: student level; Level-2: school level.

#### *Baseline Model and The Intra-class Correlation Coefficient*

First of all, a baseline model which does not contain any independent variables was run to determine how the variance components were partitioned among the two levels. Model 0 was displayed in equation below where  $i$  represents students,  $j$  represents schools,  $r_{ij}$  represents level-1 residuals,  $u_{0j}$  represents level-2 residuals,  $\beta_{0j}$  represents random effects, and  $\gamma_{00}$  represents fixed effects.

$$PV1SKCO_{ij} = \beta_{0j} + r_{ij} \text{ where } \beta_{0j} = \gamma_{00} + u_{0j}$$

The intra-class correlation coefficient (ICC;  $\rho$ ) representing the proportion of the variance accounted by the grouping variable (school membership) was calculated. The

*Investigation of Turkey's PISA 2015 Science Performance and Associated Variables Using Hierarchical Linear Modeling*

calculation of ICC is as displayed below where  $\tau_{00}$  represents level-2 variance whereas  $\sigma^2$  represents level-1 variance.

$$\rho = \frac{\tau_{00}}{\tau_{00} + \sigma^2}$$

*Additional Variance Explained*

In a two-level HLM, variance is partitioned among the two levels; therefore, the amount of variance explained at each level needs to be calculated. The equations below, offered by Snijder and Bosker (1999), display how the additional variance is explained at each level. The  $R_{L1}^2$  and  $R_{L2}^2$  represent the proportion of the amount of variance explained at level-1 and level-2, respectively. In the equations,  $\sigma^2$  represents within class variance,  $\tau$  represents between class variance,  $m_0$  represents baseline model,  $m[i]$  represents the model for which the amount of additional explained variance to be calculated,  $B$  stands for the average school sample size.

$$R_{L1}^2 = 1 - \frac{\sigma_{m[i]}^2 + \tau_{m[i]}}{\sigma_{m_0}^2 + \tau_{m_0}}; R_{L2}^2 = 1 - \frac{\sigma_{m[i]/B}^2 + \tau_{m[i]}}{\sigma_{m_0/B}^2 + \tau_{m_0}}$$

In Model 1, the relationship between SCIENCE and affective characteristics towards science is examined. Model 1 had level-1 predictors with fixed slopes and random intercepts. In other words, it was assumed that the average science performance level differed from school to school but the magnitude of the relationship between science performance and affective characteristics towards science remained constant among the schools.

$$SCIENCE_{ij} = \beta_{0j} + \beta_{1j}(EPIST) + \beta_{2j}(JOYSCI) + \beta_{3j}(SCIEEFF) + \beta_{4j}(INTBRSCI) + \beta_{5j}(MOTIVAT) + \beta_{6j}(INTSCIE) + r_{ij}$$

$$\beta_{0j} = \gamma_{00} + u_{0j}; \beta_{1j} = \gamma_{10}; \beta_{2j} = \gamma_{20}; \beta_{3j} = \gamma_{30}; \beta_{4j} = \gamma_{40}; \beta_{5j} = \gamma_{50}; \beta_{6j} = \gamma_{60}$$

Model 2 examines the relationship between SCIENCE and students' perceptions on learning environment, teacher, and teaching method. All of the variables used in this model were student level variables with the following labels DISCLISCI, IBTEACH, TEACHSUP, TDTEACH, PEFEED, ADINST, and SCIEACT. This model is also a random intercepts model.

$$SCIENCE_{ij} = \beta_{0j} + \beta_{1j}(DISLISCI) + \beta_{2j}(IBTEACH) + \beta_{3j}(TEACHSUP) + \beta_{4j}(TDTEACH) + \beta_{5j}(PEFEED) + \beta_{6j}(ADINST) + \beta_{7j}(SCIEACT) + r_{ij}$$

$$\beta_{0j} = \gamma_{00} + u_{0j}\beta_{1j} = \gamma_{10}; \beta_{2j} = \gamma_{20}; \beta_{3j} = \gamma_{30}; \beta_{4j} = \gamma_{40}; \beta_{5j} = \gamma_{50}; \beta_{6j} = \gamma_{60}; \beta_{7j} = \gamma_{70}$$

Model 3 was used to investigate the relationship between SCIENCE and the demographic characteristics of students. Just like the previous models, this was also a random intercepts model. GRADE, GENDER, SES, SCTYPE, and TRATIO are used as independent variables. GRADE, GENDER, and SES were level-1 variables whereas SCTYPE and TRATIO were level-2 variables.

$$SCIENCE_{ij} = \beta_{0j} + \beta_{1j}(GRADE) + \beta_{2j}(GENDER) + \beta_{3j}(SES) + r_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(SCTYPE) + \gamma_{02}(TRATIO) + u_{0j}\beta_{1j} = \gamma_{10}; \beta_{2j} = \gamma_{20}; \beta_{3j} = \gamma_{30}$$

Model 4 was the model that included all of the independent variables used in the previous 3 models. There are a total of 16 independent variables included. This model was also a random intercepts model as the previous models.

$$SCIENCE_{ij} = \beta_{0j} + \beta_{1j}(EPIST) + \beta_{2j}(JOYSCI) + \beta_{3j}(SCIEEFF) + \beta_{4j}(INTBRSCI) + \beta_{5j}(MOTIVAT) + \beta_{6j}(DISLISCI) + \beta_{7j}(IBTEACH) + \beta_{8j}(TEACHSUP) + \beta_{9j}(TDTEACH) + \beta_{10j}(PEFEED) + \beta_{11j}(ADINST) + \beta_{12j}(INSTSCIE) + \beta_{13j}(SCIEACT) + \beta_{14j}(GRADE) + \beta_{15j}(GENDER) + \beta_{16j}(SES) + r_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(SCTYPE) + \gamma_{02}(TRATIO) + u_{0j}\beta_{1j} = \gamma_{10}\beta_{1j} = \gamma_{10}; \beta_{2j} = \gamma_{20}; \beta_{3j} = \gamma_{30}; \beta_{4j} = \gamma_{40}; \beta_{5j} = \gamma_{50}; \beta_{6j} = \gamma_{60}; \beta_{7j} = \gamma_{70}; \beta_{8j} = \gamma_{80}; \beta_{9j} = \gamma_{90}; \beta_{10j} = \gamma_{100}; \beta_{11j} = \gamma_{110}; \beta_{12j} = \gamma_{120}; \beta_{13j} = \gamma_{130}; \beta_{14j} = \gamma_{140}; \beta_{15j} = \gamma_{150}; \beta_{16j} = \gamma_{160};$$

#### *Model-data Fit and Comparison of Models*

Some of the very common comparative fit indices that were used to compare a variety of nested models are AIC (Akaike, 1987), BIC (Bozdogan, 1987) and DIC. All of these statistics are approximations of chi-square model-data fit index that could be used to see if there was any improvement in model-data fit when comparing a variety of nested models. Smaller values of AIC, BIC, DIC are indication of the improvement in model-data fit. Since these three statistics are not absolute, chi-square likelihood-ratio tests for comparing the nested models will be conducted in order to see if additional parameters improve the fit significantly.

*Effect Sizes*

The parameters estimates were converted to effect sizes so that the relative importance of the variables could be judged. In order to achieve that goal, a statistics equivalent to Cohen's  $d$  pointed out by Spybrook, Raudenbush, Liu, Congdon, and Martínez (2006) was used. The formula used to calculate the effect sizes is below:

$$\delta = \frac{\hat{\gamma}_{01}}{\sqrt{\hat{\sigma}^2 + \hat{\tau}_{00}}}$$

*Estimation and Software*

Maximum-likelihood was used as the method of estimation due to its flexibility offered in comparing nested models. Another property that makes maximum-likelihood preferable is that it can produce more precise results when there is great flexibility in the sample size of the higher levels (Albright & Marinova, 2010). The data was analyzed using R software (R Core Team, 2018) version 1.1.4402 and Mplus 6 (Muthen & Muthen, 2007). The R packages that were used to run the analyses were lme4 (Bates, Mächler, Bolker & Walker, 2014), Ggplot2 (Wickham, 2016) and Naniar (Tierney, Cook, McBain & Fay, 2018), and mitml (Grund, Robitzsch & Luedtke, 2018).

**Findings***Handling Missing Data*

The dependent variable (SCIENCE) and the demographic variables did not have any missing values; however, independent variables had missing data by the amount varying from a variable to another one. Figure 1 displays the pattern of missingness as it appears in the data. There were 37 cases that had missing values on the type of school they were attending. These cases were removed from the dataset. After the screening for missing data and removal of the cases mentioned above, the sample size reduced from 5581 to 5544. A series of analyses were performed to determine the nature of missingness, and the data was found to be missing at random (MAR; Enders, 2010). The mechanism of the missing data was dependent on science performance (SCIENCE) and socio-economic status (SES). Therefore, a joint multiple imputation approach, implemented in the R package 'mitml' (Grund, Robitzsch, Luadtke, 2019), was applied which involved both variables containing missing values as well as the plausible values (SCIENCE, SES etc.) provided in PISA. Imputation of missing data was a little complex since there were 10 plausible values for SCIENCE, and 10 plausible

values for SES. Using the 10 pairs (SCIENCE and SES) of plausible values, 5 imputation datasets for each pair were generated using mitml (Grund, Robitzsch & Luedtke, 2018) implemented in R. As a result, a total of 50 datasets were generated for further analysis. The imputation method was a model-based method that used a two level HLM. The estimation of missing observation was done via Bayesian approach using 500 burn-in iterations followed by 5000 actual iterations.

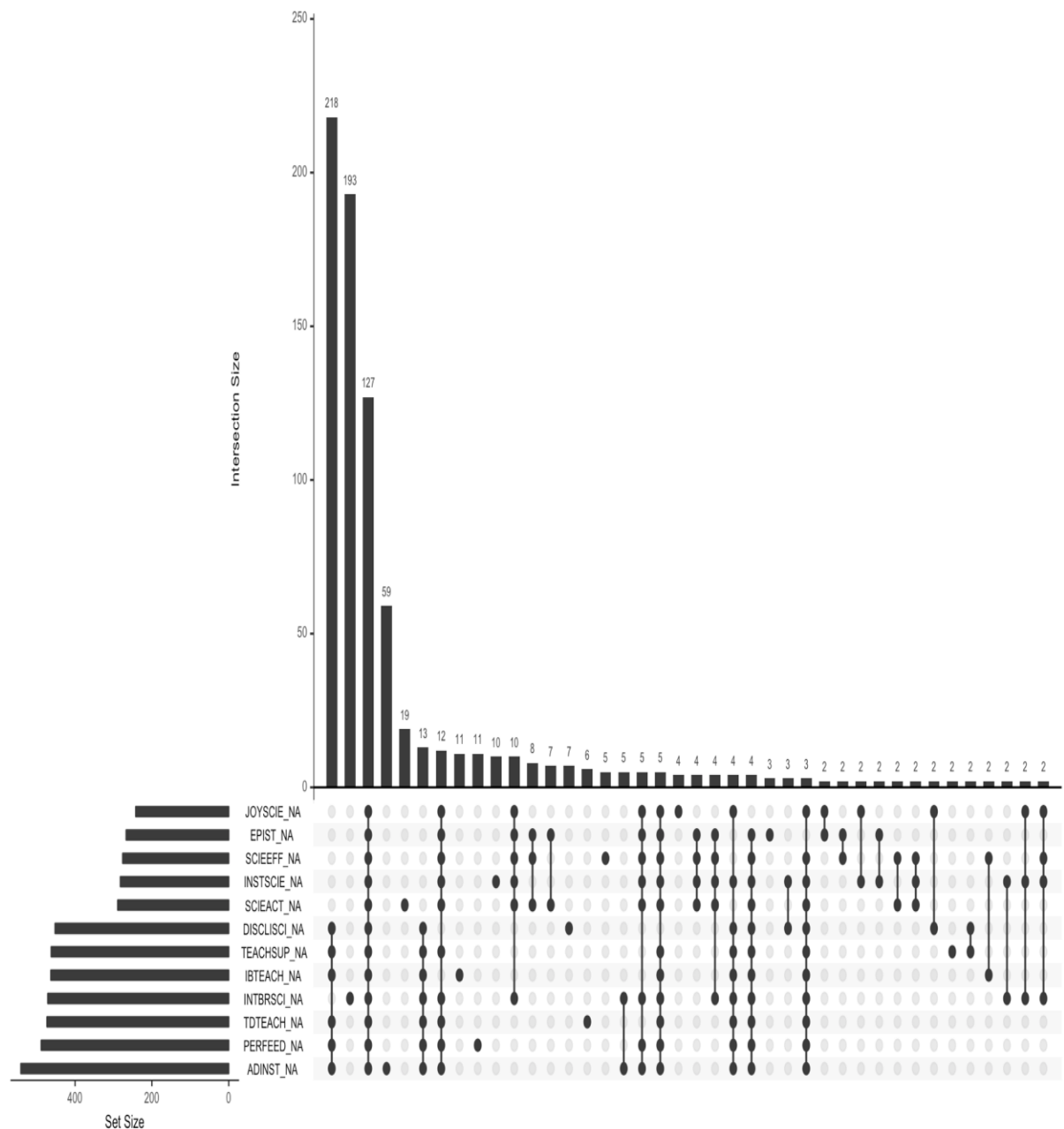


Figure 1 The Pattern of Missing Data Across Independent Variables



*Investigation of Turkey's PISA 2015 Science Performance and Associated Variables Using Hierarchical Linear Modeling*

The imputed datasets having the plausible values, were analyzed using Mplus version 6 (Muthen & Muthen, 2007). Table 5 displays the fixed parameter estimates, their statistical significance, effect sizes, random effects, the percentage of the additional variance explained, and the model-data fit statistics. All of the values displayed in the table are a summary of the 50 imputed datasets.



Araştırma Makalesi / Research Article

**Table 5** Parameter Estimates, Standard Errors, Significances, Fixed and Random Effects, and Model-Data Fit

Fixed effects:																				
	Model 0				Model 1				Model 2				Model 3				Model 4			
	Est.	SE	t	p	Est.	SE	Eta	p	Est.	SE	Eta	p	Est.	SE	Eta	p	Est.	SE	Eta	p
Intercept	<b>419</b>	<b>4.54</b>	<b>92</b>	*	<b>419</b>	<b>4.3</b>	-	*	<b>419</b>	<b>4.43</b>	-	*	<b>441</b>	<b>8.30</b>	-	*	<b>438</b>	<b>6.86</b>	-	*
EPIST	-	-	-	-	<b>6.24</b>	<b>1.06</b>	<b>0.08</b>	*	-	-	-	-	-	-	-	-	1.31	1.09	0.03	0.23
JOYSCI	-	-	-	-	<b>5.81</b>	<b>0.97</b>	<b>0.07</b>	*	-	-	-	-	-	-	-	-	<b>4.27</b>	<b>1.08</b>	<b>0.09</b>	*
SCIEEFF	-	-	-	-	<b>4.44</b>	<b>1.05</b>	<b>0.06</b>	*	-	-	-	-	-	-	-	-	<b>3.55</b>	<b>0.97</b>	<b>0.07</b>	*
MOTIVAT	-	-	-	-	1.29	1.25	0.02	0.30	-	-	-	-	-	-	-	-	-0.32	1.26	-0.00	0.79
INTBRSCI	-	-	-	-	<b>2.93</b>	<b>1.11</b>	<b>0.04</b>	*	-	-	-	-	-	-	-	-	<b>2.04</b>	<b>0.96</b>	<b>0.04</b>	*
INTSCIE	-	-	-	-	-1.1	1.17	-0.0	0.33	-	-	-	-	-	-	-	-	0.60	1.54	0.01	0.69
DISCLISCN	-	-	-	-	-	-	-	-	1.75	1.00	0.02	0.08	-	-	-	-	-0.36	1.07	-0.00	0.73
IBTEACH	-	-	-	-	-	-	-	-	<b>-7.21</b>	<b>1.02</b>	<b>-0.09</b>	*	-	-	-	-	<b>-2.56</b>	<b>1.04</b>	<b>-0.05</b>	*
TEACHSUP	-	-	-	-	-	-	-	-	<b>2.39</b>	<b>1.18</b>	<b>0.03</b>	*	-	-	-	-	0.17	1.34	0.00	0.89
TDTEACH	-	-	-	-	-	-	-	-	<b>2.96</b>	<b>1.12</b>	<b>0.04</b>	*	-	-	-	-	1.45	1.37	0.03	0.29
PERFEED	-	-	-	-	-	-	-	-	<b>-3.16</b>	<b>1.17</b>	<b>-0.04</b>	*	-	-	-	-	-1.34	1.20	-0.03	0.26
ADINST	-	-	-	-	-	-	-	-	<b>4.19</b>	<b>1.39</b>	<b>0.05</b>	*	-	-	-	-	-0.43	1.19	-0.00	0.71
SCIEACT	-	-	-	-	-	-	-	-	0.29	1.01	0.00	0.77	-	-	-	-	-1.20	1.16	-0.02	0.30
GENDER	-	-	-	-	-	-	-	-	-	-	-	-	<b>7.63</b>	<b>1.91</b>	<b>0.15</b>	*	<b>8.24</b>	<b>1.93</b>	<b>0.17</b>	*
SES	-	-	-	-	-	-	-	-	-	-	-	-	<b>54.59</b>	<b>1.43</b>	<b>1.12</b>	*	<b>52.94</b>	<b>1.52</b>	<b>1.11</b>	*
GRADE	-	-	-	-	-	-	-	-	-	-	-	-	<b>-7.71</b>	<b>2.30</b>	<b>-0.16</b>	*	<b>-7.78</b>	<b>2.30</b>	<b>-0.16</b>	*
TRATIO	-	-	-	-	-	-	-	-	-	-	-	-	1.04	2.31	0.00	0.65	0.70	2.31	0.00	0.75
SCTYP	-	-	-	-	-	-	-	-	-	-	-	-	-14.23	8.96	-0.28	0.11	-10.88	7.08	-0.27	0.12
Random effects:																				
	Model0				Model1				Model2				Model3				Model4			
Residual $r_{ij}$	3388				3242				3323				1845				1789			
Intercept $u_{0j}$	3367				2796				2883				509				486			

**Explained additional variance:**

	Model0	Model1	Model2	Model3	Model4
Additional $R_{L1}^2$		0.06	0.03	0.63	0.64
Additional $R_{L2}^2$		0.08	0.05	0.82	0.82

**Model-data fit\*\*:**

	Model0	Model1	Model2	Model3	Model4
AIC	61347(80)	61106(91)	61249(93)	67462(70)	67313(71)
BIC	61367(80)	61165(91)	61315(93)	67528(70)	67465(71)
DIC	61338(78)	61085(90)	61226(92)	57794(71)	57619(71)

\*Statistically significant finding at alpha=0.05 finding;\*\* Values inside the parenthesis are standard deviations of the estimates across 50 imputed datasets. SE: Standard error of the fixed effect; Bold text: statistically significant finding at alpha=0.05; Intercept:  $u_{0j}$  = the residual of the intercept for school j; Residual:  $r_{ij}$ = the residual for  $i^{\text{th}}$  student and  $j^{\text{th}}$  school;  $R_{L1}^2$ = the proportion of the additional variance explained at level-1 when predictors were added to the empty model.  $R_{L2}^2$ = the proportion of the additional variance explained at level-2 when predictors were added to the empty model; AIC: Akaike Information Criteria; BIC: Bayesian Information Criteria; DIC: Deviance Information Criteria; loglik: log-likelihood; df: degrees of freedom. Eta: relative importance of a predictor in a model in the form of a z-score.



Araştırma Makalesi / Research Article

The intra-class correlation coefficient was as 0.47 using Model 0, which would be considered as large. This size of an ICC justifies the use of an HLM model given the structure of the data (Cohen, Cohen, West & Aiken, 2003). The meaning that could be attached to this ICC is that 47% of the variability found in the science performance is due to the school students attend. In other words, there is a great amount of performance gaps among the schools, which can explain 48% of the variability in SCIENCE. The additional amount of variance explained by Model 1 on top of Model 0 were 6% and 8% for level-1 and level-2, respectively. Further, the model-data fit statistics (AIC, BIC, DIC) of Model 1 showed improvement compared to Model 0. As displayed in Table 6, EPIST had the largest effect size (0.08) implying that it is the strongest predictor of science performance among the independent variables listed for Model 1.

Model 2 had 7 independent variables that were used to investigate the relationship between learning environment and science performance. Two of the 7 independent variables, SCIEACT and DISCLISCN, did not have statistically significant slope estimates. The inclusion of these 7 independent variables explained 3% and 5% additional variance of science performance at level-1 and level-2, respectively. The model-data fit statistics showed that Model 2 fits better to the data than Model 0. In terms of effect sizes, ADINST was the strongest predictor of science performance for Model 2 ( $\delta_{ADINST} = 0.05$ ).

Model 3 was designed to investigate the relationship between science performance and demographic variables. Two of the fixed parameter estimates, TRATIO and SCTYP, were not statistically significant. The amount of additional variance in science performance explained by the demographics was 63% and 82% for level-1 and level-2, respectively. This model had a huge improvement in terms of model-data fit compared to Model 0. SES was the strongest predictor of science performance in this model with a very large effect size of 1.12 which was followed by SCTYP, that had an effect size of -0.28 favoring students who attend to the private schools.

Lastly, Model 4 was run with all of the independent variables used in the previous 3 models. This model was able to explain 64% and 82% additional variance of science performance at level-1 and level-2, respectively. This model had 7 out of 18 independent

variables that were statistically significant. This model had better model-data fit than all other models based on the model-data fit statistics displayed in Table 5.

Chi-square ( $X^2$ ) difference tests were run to compare the nested models in terms of model-data fit. Only models that are nested could be compared using a chi-square difference test. In other words, Model A is considered to be nested in Model B only if Model B includes all of the parameters that Model A has. Table 6 displays the chi-square difference tests that were run to compare the models that were nested.

**Table 6** Comparison The Models That are Nested

Model A	Model B	$X^2_{difference}$	df	p
Model 4	Model 0	3718.72	18	<0.05***
Model 4	Model 1	3465.92	13	<0.05***
Model 4	Model 2	3607.01	12	<0.05***
Model 4	Model 3	175.38	11	<0.05***
Model 1	Model 0	252.90	5	<0.05***
Model 2	Model 0	111.81	6	<0.05***
Model 3	Model 0	3543.44	7	<0.05***

\*\*\*Significant at 0.05; df: degrees of freedom

The findings show that Model 4 has explained more variance than each of the remaining models as displayed in Table 5. The chi-square difference tests, displayed in Table 6, comparing Model 4 to the other 4 models showed that the fit of Model 4 to the data outperformed all the other models. In addition, Model 0 is nested and explained less variance than each of the remaining models. The chi-square difference tests indicated that each of the non-empty models (Model 1 through Model 4) explained significantly more amount of variance than Model 0. What these findings imply is that each of the non-empty models explained certain number of additional variances that cannot be ignored. Since Model 1, Model 2, and Model 3 are not nested in one another, it is impossible to compare them to each other using statistical significance tests. Their fit to the data was approximately examined using additional variances explained at level-1 and level-2. Using the additional variances explained, Model 3 was selected as the best model among the three models since it explained 63% and 82% additional variance at level-1 and level-2, respectively. The additional variances explained by Model 1 were higher than Model 2 as displayed in Table 5.

## Result and Discussion

In this study, which examines the relationship between Turkish 9th and 10th grade

students' science performance, and a set of variables that are related to learning environment, affective characteristics, and demographics, it was realized that variables in Model 3, student demographic characteristics (GENDER, SCTYP, SES, GRADE and TRATIO), are strongly related to science performance explaining 63% and 83% of the unexplained variance. Although Model 4 has 13 more independent variables than Model 3, Model 4 was able to explain only 1% additional variance, which is not a remarkable contribution. These findings revealed that a large part of the differences among students' science performances could be explained by Model 3 (demographic characteristics of students). Except for two of the variables in this model (TRATIO and SCTYP), the others (SES, GENDER, and GRADE) were found to be statistically related with the students' science performance. As a result, it could be concluded that science performance in Turkey primarily effected by demographic characteristics. Again, it can be said that the attitudes and beliefs of the students towards science effect their science performance more than the characteristics of the learning environment. However, their total contribution is way less than student demographic characteristics.

It was observed that the effect size of the SES (1.12), was clearly ahead of the other variables. This finding is consistent with other studies in the literature. For example, in two different studies examining the PISA 2006 data of 25 different countries (Australia, Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, South Korea, Luxembourg, the Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, the UK and the USA), it was found that the socioeconomic status of the students was significantly associated with science performance (Alivernini & Manganelli, 2015; Sun et al. 2012).

Another variable that had an effect size close to the SES was the SCTYP variable, which represents whether the school is a private school or not. The effect of SCTYP reflects the difference between students studying in private schools and the students in public schools. The difference between the public schools and the private schools was not statistically significant most probably due to low sample size and power issues.. In a study that examined the PISA 2009 dataset in Singapore, Areepattamannil et al. (2015) found that the socioeconomic level of a school was positively related to science performance while school type (public or private), number of students per teacher, and quality of the educational resources of the school did not. Turkey sample compared with a sample of Australia have similar characteristics in this respect. Perry and McConney (2004) reported that students with

high socioeconomic status prefer attending private schools in Australia; on the contrary, students with low socioeconomic status usually have public school option only. Private schools in Australia have two sources of funding: the state and the parents while public schools only meet their needs with the funds they receive from the state. Private schools meet their needs with the fees they receive during registration, and they use these funds to increase the quality of educational resources. For this reason, public school versus private school achievement gap in Australia could be due to educational resources (Perry & McConney, 2004). Again, these researchers reported that the educational programs implemented in public and private schools were not the same since more intense focus is on academic preparations for university entrance exams, and strict academic curriculum orientation in these schools might explain the high scores in PISA. There is a similar situation in Turkey that schools vary on their average socioeconomic status. The effect of the type of school on students' science performance in Turkey may also be due to similar reasons. In contrast, in a study evaluating the PISA 2006 and 2009 data of 10 OECD countries, it was pointed out that school type made a difference on student performance in favor of public school student, except Australia (Sousa et al. 2012). It can be inferred that in countries where the standards of private and public schools are the same, the school type does not affect the performances of students. In a study with Australian data, it was found that the socioeconomic level of the school was related to the students' science performance levels, and that this relationship was similar for all students regardless of individual socioeconomic levels (Perry & McConney, 2004). Based on this, it is thought that, if the differences between the schools are reduced and the socioeconomic levels of the public schools are increased, the students' performances can be increased independently of their individual socioeconomic levels. Lastly, OECD 2004 report states that more inclusive school systems (e.g., Canada and Finland), in which there is no difference between types of school will both be effective in achieving higher performance levels, and that students will have less inequality due to different socioeconomic status compared to more less inclusive school systems (e.g., Australia and Turkey) despite there are more schools (OECD, 2004).

In terms of effect sizes, SES and SCTYP variables had largest effects followed by GRADE and GENDER variables, respectively. The research findings in this study revealed that students at the 10<sup>th</sup> grade were more successful than the 9<sup>th</sup> grade students. This is thought to be due to the fact that the 10<sup>th</sup> grade students could be more matured on their science achievement. Further, research findings in this study revealed that male students were more successful than female students. When the literature was examined, it was found that

these findings were not consistent with the findings of other studies.-In previous PISA studies on Turkey sample which examined the relationship between the science performance and gender, it was reported that gender variable to be positively associated with success in favor of female students (Akkuş, 2008; Gürsakal, 2012). The ratio of females attending to school from socioeconomically disadvantaged families in Turkey is known to be low (Kocabaş Aladag & Yavuzalp, 2004). Therefore, in PISA 2015, the same phenomena could have led the girls to score less than the boys.

There were 6 independent variables tested on Model 1. The regression slope estimates were statistically significant except for MOTIVAT and INSTSCIE. EPIST and JOYSCI variables follow GENDER in terms of their effect on student science performance (see Table 5). Our findings regarding the effect of epistemological beliefs on student performance are consistent with the findings in the relevant literature (Mason, et al. 2013; Topçu & Yılmaz Tüzün, 2009). Our findings point out that students' epistemological beliefs can guide their own scientific knowledge acquisition and give shape to science-learning orientations (Tsai, 2006). This means that students' epistemological beliefs should be improved in order to increase their science performance. However, the effect of EPIST on science performance was 0.08, which is considered as a small effect size. In addition, the finding that the enjoyment of science variable (JOYSCI) was significantly associated with student performance was consistent with the field literature (Lam & Lau, 2014; Yetişir et al. 2018). The findings of the studies in the literature in parallel with our findings suggest that the self-efficacy variable (SCIEEFF) is positively related to student performance (Alivernini & Manganelli, 2015; Lam & Lau, 2014). It is expected that the students who have higher-level efficacy could be more successful. Research findings revealed that instrumental motivation towards science (INSTSCIE) was not significantly associated with student performance.

Our findings revealed that motivation towards science (MOTIVAT) was not statistically significant predictor of science performance. Findings on motivation in this study are incompatible with the PISA 2006 Hong Kong sample (Sun et al. 2012). Student interest and motivation towards science in Turkey seem to be higher than the OECD average level. They find themselves more adequate (high-self efficacy) in this area than average of the other OECD countries. It was also seen that the ratio of students expecting to have a profession related to science is higher than the OECD average (Taş, Arıcı, Özarkan & Özgürlük, 2016). However, when examining Turkey PISA 2015 results related to science performance tests, although students' interest and motivation towards science is high, it is noteworthy that they



remain below the OECD average in science performance tests. This situation supports our finding that the demographic variables (especially socioeconomic level, school, public school or private school) have stronger effect on student science performance than the affective characteristics. Overwhelming influence of variables caused by the students themselves in Turkey sample shows that the students of the high interest and motivation alone is not sufficient for them to be successful on their science.

Six of the variables used in Model 2 measuring the instructional environment characteristics (DISCLISCN, TEACHSUP, TDTEACH, ADINST, IBTEACH, PERFEED) were found to be significant predictors of student performance. The effect sizes of these variables were found to be lower than the effect sizes of demographic characteristics and affective characteristics. The fact that the degree of disciplinary nature of the classroom environment (DISCLISCN) is significantly associated with student performance is consistent with Sousa et al. (2012)'s the findings that evaluated PISA 2006 and 2009 datasets. These findings indicated that increasing the level of discipline in the classroom will increase science achievement is a myth because the contents high success rate of countries with low disciplinary levels such as Finland and Singapore refute the idea .

In addition, this study has found an interesting finding on the relationship between the variable related to 'how the teaching is done' and science performance. While the expectation was exact opposite, it was found that the effect of the inquiry-based science teaching and learning approach (IBTEACH) on student performance was below 0.1. That is, although the relationship between IBTEACH variable and student performance seems statistically significant, it has a low effect size. Research findings in this study are inconsistent with the findings of other studies in the field literature (Cairns & Areepattamannil, 2017; Minner et al. 2010). Jiang and McComas (2015) found that the complex effect of this variable on success was related to the level of clarity of the inquiry-based teaching used. The reason why this variable has a low level of explaining student science performances may be related to the extent to which and how inquiry-based teaching is handled in the classroom. Although the level of perception related to the needs to be re-arranged according to needs (ADINST) and student's feedback level (PERFEED) are positively related to the student performance, the effect of these variables on explaining science performance was less than 0.1. It is thought that this situation can be caused by the lack of perceptions of the students towards to teachers' feedbacks, because when teachers provide feedback to support teaching, students use deeper learning strategies which tends to make teaching more permanent (Young, 2005).

The finding that the level of teacher support (TEACHSUP) is a positive predictor of student science performance is compatible with previous research (Brophy, 2004; Hardre & Sullivan, 2008 as cited in: Yıldırım, 2012). Finally, the fact that SCIEACT variable is not effective at predicting science performance might be due to the quality of the organized science activities. Another possible explanation is that science activities organized in the classroom may not be capable of developing high-level cognitive outcomes or improving the cognitive capacity of students (Aydoğdu, 2006; Kıncal & Yazgan, 2010).

### **Suggestions**

This study reveals important clues about new goals and reforms, which will guide to Turkish educational system. Based on these findings, it is thought that improving students' science performance is possible by taking into account the variables that effect the performance, especially with improvements that can provide equal opportunities in the context of socioeconomic status, and the construction of inclusive school systems. This can be done possibly by planning the budget allocated for education in a way that ensures the schools in regions with low socio-economic level are improved in terms of physical and educational resources. It is recommended to allocate the budget allocated for education among schools to provide inclusive school systems. It is essential for politicians and educators to act together to eliminate the difference between public and private schools in terms of physical conditions, resources, education and preparations for university entrance examinations. It is recommended that politicians and education programmers should take this into consideration. Further, gender ranked as the 3<sup>rd</sup> variable in terms of the size of its effect at 0.17 in favor of boys. In our educational system, we need to create steps for equal opportunities for different gender groups.

### **Acknowledgements**

A part of this study was presented at the International Learning, Teaching and Educational Research (ILTER) Congress, Amasya University, Turkey (2018).

### **References**

- Acar, T., & Ogretmen, T. (2012). Analysis of 2006 PISA science performance via multilevel statistical methods. *Education and Science*, 37(163), 178-189.
- Albright, J. J., & Marinova, D. M. (2010). *Estimating multilevel models using SPSS, Stata, SAS, and R*. Bloomington, IN: Indiana University.

- Akaike, H. (1987). *Factor analysis and AIC*. In *Selected Papers of Hirotugu Akaike* (pp. 371-386). Springer, New York, NY.
- Akkuş, N. (2008). *Yaşam boyu öğrenme becerilerinin göstergesi olarak 2006 PISA sonuçlarının Türkiye açısından değerlendirilmesi [Life-long learning skills as an indicator of the PISA 2006 results in terms of assessing Turkey]*. Master thesis, Hacettepe University, Ankara, Turkey.
- Alivernini, F., & Manganelli, S. (2015). Country, school and students factors associated with extreme levels of science literacy across 25 countries. *International Journal of Science Education*, 37(12), 1992-2012.
- Areepattamannil, S. (2014). International Note: What factors are associated with reading, mathematics, and science literacy of Indian adolescents? A multilevel examination. *Journal of adolescence*, 37(4), 367-372.
- Areepattamannil, S., Chiam, C. L., Lee, D. H., & Hong, H. (2015). Correlates of science achievement in Singapore: a multilevel exploration. In *Science Education in East Asia* (pp. 607-629). Springer, Cham.
- Aydoğdu, B. (2006). *İlköğretim fen ve teknoloji dersinde bilimsel süreç becerilerini etkileyen değişkenlerin belirlenmesi [Identification of variables effecting science process skills in primary science and technology course]*. Doctoral dissertation. Dokuz Eylül University, Izmir, Turkey.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48.
- Bozdoğan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52(3), 345-370.
- Buuren, S. V., & Groothuis-Oudshoorn, K. (2010). Mice: Multivariate imputation by chained equations in R. *Journal of statistical software*, 55(2), 1-68.
- Bybee, R. & McCrae, B. (2011). Scientific literacy and student attitudes: Perspectives from PISA 2006 science. *International Journal of Science Education*, 33(1), 7-26.
- Cairns, D., & Areepattamannil, S. (2017). Exploring the relations of inquiry-based teaching to science achievement and dispositions in 54 countries. *Research in Science Education*, 49(1), 1-23.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. New York, NY: Psychology Press.

- De Ayala, R. J. (2013). *The theory and practice of item response theory*. Guilford Publications.
- Enders, C. K. (2010). *Applied missing data analysis*. Guilford press.
- Finch, W. H., Bolin, J. E., & Kelley, K. (2016). *Multilevel modeling using R*. Boca Raton: Crc Press.
- Grund, S., Robitzsch, A., & Luedtke, O. (2018). Mitml: Tools for Multiple Imputation in Multilevel Modeling (2018). *R package version 0.3-6*.
- Gürsaka, S. (2012). PISA 2009 öğrenci başarı düzeylerini etkileyen faktörlerin değerlendirilmesi [An evaluation of PISA 2009 student achievement levels' affecting factors]. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 17(1), 441-452.
- Güzel, Ç. I., & Berberoğlu, G. (2005). An analysis of the Programme for International Student Assessment 2000 (PISA 2000) mathematical literacy data for Brazilian, Japanese, and Norwegian students. *Studies in Educational Evaluation*, 31, 283-314.
- Jiang, F., & McComas, W. F. (2015). The effects of inquiry teaching on student science achievement and attitudes: Evidence from propensity score analysis of PISA data. *International Journal of Science Education*, 37(3), 554-576.
- Kıncal, R. Y., & Yazgan, A. D. (2010). İlköğretim 7. ve 8. sınıf öğrencilerinin formal operasyonel düşünme becerilerinin bazı değişkenler açısından incelenmesi [Investigating the Formal Operational Thinking Skills of 7th and 8th Grade Primary School Students According to Some Variables]. *Elementary Education Online*, 9(2), 723-733<sup>[L]</sup><sub>[SEP]</sub>
- Kocabaş, İ., Aladağ, S., & Yavuzalp, N. (2004). *Eğitim sistemimizdeki okullaşma oranlarının analizi [Analysis of schooling rates in our education system]*. Paper presented at XIII. National Educational Sciences Congress, Inonu University, Malatya.
- Lam, T. Y. P., & Lau, K. C. (2014). Examining factors affecting science achievement of Hong Kong in PISA 2006 using hierarchical linear modeling. *International Journal of Science Education*, 36(15), 2463-2480.
- Mason, L., Boscolo, P., Tornatora, M. C., & Ronconi, L. (2013). Besides knowledge: A cross-sectional study on the relations between epistemic beliefs, achievement goals, self-beliefs, and achievement in science. *Instructional Science*, 41(1), 49-79.
- Ministry of National Education (2003). *PISA 2003 ulusal rapor [PISA 2003 National Report]*. Retrieved from PISA Turkey web site: [http://pisa.meb.gov.tr/?page\\_id=22](http://pisa.meb.gov.tr/?page_id=22)

- Milli Eğitim Bakanlığı Talim Terbiye Kurulu Başkanlığı, MEB (2005). *İlköğretim fen ve teknoloji dersi öğretim programı ve kılavuzu [Primary science and technology course curriculum and instruction]*. Ankara.
- Ministry of National Education (2006). *PISA 2006 ulusal rapor [PISA 2006 national Report]*. Retrieved from PISA Turkey web site: [http://pisa.meb.gov.tr/?page\\_id=22](http://pisa.meb.gov.tr/?page_id=22)
- Ministry of National Education (2009). *PISA 2009 ulusal rapor [PISA 2009 National Report]*. Retrieved from PISA Turkey web site: [http://pisa.meb.gov.tr/?page\\_id=22](http://pisa.meb.gov.tr/?page_id=22)
- Ministry of National Education (2012). *PISA 2012 ulusal rapor [PISA 2012 National Report]*. Retrieved from PISA Turkey web site: [http://pisa.meb.gov.tr/?page\\_id=22](http://pisa.meb.gov.tr/?page_id=22)
- Ministry of National Education (2013). *İlköğretim kurumları fen bilimleri dersi öğretim programı [Primary school institutions science courses curriculum]*. Ankara.
- Ministry of National Education (2018). *İlköğretim kurumları fen bilimleri dersi öğretim programı [Primary school institutions science courses curriculum]*. Ankara.
- Ministry of National Education (2016). *PISA 2015 projesi: Ulusal ön rapor [PISA 2015 project: National pre-report]*. Retrieved from Ankara. [http://pisa.meb.gov.tr/wp-content/uploads/2016/12/PISA2015\\_Ulusal\\_Rapor.pdf](http://pisa.meb.gov.tr/wp-content/uploads/2016/12/PISA2015_Ulusal_Rapor.pdf)
- Minner, D. D., Levy, A. J., & Century, J. (2010). Inquiry-based science instruction-what is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching*, 47(4), 474–496.
- Muis, K. R., Bendixen, L. D., & Haerle, F. C. (2006). Domain-general and domain-specificity in personal epistemology research: Philosophical and empirical reflections in the development of a theoretical framework. *Educational Psychology Review*, 18(1), 3-54.
- Muthén, L. K., & Muthén, B. O. (2007). *Statistical analysis with latent variables using Mplus*. Los Angeles, CA: Muthén & Muthén.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *ETS Research Report Series*, 1992(1), 1-30.
- Ning, B., Van Damme, J., Van Den Noortgate, W., Yang, X., & Gielen, S. (2015). The influence of classroom disciplinary climate of schools on reading achievement: A cross-country comparative study. *School Effectiveness and School Improvement*, 26(4), 586-611.
- Next Generation Science Standards (2013). *Next generation science standards: for states, by states*. Washington: National Academies Press.

- Organization for Economic Co-operation and Development (2015a). *Scaling procedures and construct validation of context questionnaire data*. Retrieved from <http://www.oecd.org>
- Organization for Economic Co-operation and Development (2015b). *Sample design*. Retrieved from <http://www.oecd.org>
- Özdemir, O. (2010). Fen ve teknoloji öğretmen adaylarının fen okuryazarlığının durumu [The status of science and technology teacher candidates' science literacy]. *Türk Fen Eğitimi Dergisi*, 7(3), 42-56.
- Özdemir, C. (2017). OECD PISA Türkiye verisi kullanılarak yapılan araştırmaların metodolojik taraması [A methodological review of research using oecd pisa Turkey data]. *Eğitim Bilim Toplum*, 14(56), 10-27.
- Perry, L. B., & McConney, A. (2010). Does the SES of the school matter? An examination of socioeconomic status and student achievement using PISA 2003. *Teachers College Record*, 112(4), 1137-1162.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.
- R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.R-project.org/>.
- Ryder, J. (2001). Identifying science understanding for functional scientific literacy. *Studies in Science Education*, 36, 1-42.
- Sadıç, A., & Çam, A. (2015). Eight grade students' epistemological beliefs with pisa success and their scientific literacy. *Journal of Computer and Education Research*, 3(5), 18-49.
- Shin, J., Lee, H., & Kim, Y. (2009). Student and school factors affecting mathematics achievement: International comparisons between Korea, Japan and the USA. *School Psychology International*, 30(5), 520-537.
- Snijders, T., & Bosker, R. (1999). *Multilevel modeling: An introduction to basic and advanced multilevel modeling*. London, Sage.
- Sousa, S., Park, E. J., & Armor, D. J. (2012). Comparing effects of family and school factors on cross-national academic achievement using the 2009 and 2006 PISA surveys. *Journal of Comparative Policy Analysis: Research and Practice*, 14(5), 449-468.

- Spybrook J, Raudenbush SW, Liu X, & Congdon R. (2006). *Optimal design for longitudinal and multilevel research: Documentation for the "Optimal Design" software*. University of Michigan, Ann Arbor, MI. [[Google Scholar](#)]
- Sun, L., Bradley, K. D., & Akers, K. (2012). A multilevel modelling approach to investigating factors impacting science achievement for secondary school students: PISA Hong Kong sample. *International Journal of Science Education*, 34(14), 2107-2125.
- Taş, U. E., Arıcı, Ö., Ozarkan, H. B., & Özgürlük, B. (2016). *PISA 2015 ulusal raporu. [PISA 2015 national report]* Ankara: Milli Eğitim Bakanlığı.
- Tierney, N., Cook, D, McBain, M., & Fay, C. (2018). Naniar: Data Structures, Summaries, and Visualisations for Missing Data. R package version 0.4.0.0. Retrieved from <https://CRAN.R-project.org/package=naniar>
- Topçu, M. S., & Yılmaz Tüzün, Ö. (2009). Elementary students' metacognition and epistemological beliefs considering science achievement, gender and socioeconomic status. *Elementary Education Online*, 8(3), 676-693<sup>[11]</sup><sub>SEP</sub>.
- Tsai, C. (2006). Reinterpreting and reconstructing science: Teachers' view changes towards the nature of science by courses of science education. *Teaching and Teacher Education*, 22(3), 363-375.
- Weirich, S., Haag, N., Hecht, M., Böhme, K., Siegle, T., & Lüdtke, O. (2014). Nested multiple imputation in large-scale assessments. *Large-scale assessments in education*, 2(1), 9.
- Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. New York. NY: Springer.
- Wigfield, A., & Eccles, J.S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, 25(1), 68-81.
- Wise, K. C., & Okey, J. R. (1983). A meta-analysis of the effects of various science teaching strategies on achievement. *Journal of Research in Science Teaching*, 20(5), 419-435.
- Yetişir, M. İ., Batı, K., Kahyaoğlu, M., & Birel, F. K. (2018). Investigation of the relation of disadvantaged students to affective characteristics of science literacy performances]. *Ankara University Journal of Faculty of Educational Sciences*, 51(1), 143-158.
- Yıldırım, S. (2012). Teacher support, motivation, learning strategy use, and achievement: A multilevel mediation model. *The Journal of Experimental Education*, 80(2), 150-172.
- Young, M. R. (2005). The motivational effects of the classroom environment in facilitating self-regulated learning. *Journal of Marketing Education*, 27, 25-4.