



## Fen Başarısı Açısından Okullar Arası Farklar: TIMSS 2019

Metehan GÜNGÖR<sup>1\*</sup> ve Sinan M. BEKMEZCİ<sup>2</sup>

<sup>1</sup>Eğitimde Ölçme ve Değerlendirme / Eğitim Bilimleri Enstitüsü, Ankara Üniversitesi, TÜRKİYE

<sup>2</sup>Matematik ve Fen Bilimleri Eğitimi / Fen Bilimleri Enstitüsü, Manisa Celal Bayar Üniversitesi, TÜRKİYE

\*([gungormetehan@gmail.com](mailto:gungormetehan@gmail.com)) Başlıca yazarın mail adresi

**Özet** – Geniş ölçekli testlerin verileri genel olarak hiyerarşik yapıya sahiptir. Hiyerarşik yapıya sahip verilerin analizinde ise çok düzeyli modellemeler sıklıkla tercih edilmektedir. Bu çalışmada TIMSS’in 2019 döngüsüne 8. sınıf düzeyinde katılan tüm ülkelerin verileri kullanılarak öğrencilerin fen başarı puanlarındaki varyansın ne kadarlık kısmının okullar arası farklılıktan kaynaklandığının hesaplanması amaçlanmıştır. Araştırmanın örnekleme katılımcı 39 ülkedeki 7363 okuldaki toplam 227345 8. sınıf öğrencisinden oluşmaktadır. Veri analizleri için birinci düzeyde öğrencilerin, ikinci düzeyde okulların yer aldığı iki düzeyli hiyerarşik doğrusal modelleme (HLM) kullanılmıştır. Çıktı değişkeni (bağımlı) olarak TIMSS 2019’da raporlanan beş olası değer (plausible value) birlikte kullanılmış ve her iki düzeyde (öğrenci ve okul) de ağırlıklandırma yapılmıştır. Araştırmanın bulgularına göre fen başarı puanlarındaki varyansın okullar arasındaki farklılıktan kaynaklanan kısmı (yüzde olarak) farklı ülkelerde %6,3 (Kore) ile %58,29 (ABD) arasında değişmektedir. AB üyesi ülkelerde bu oran ortalama olarak %23,6 civarında iken, kuzey ülkelerinde (Finlandiya, İsveç ve Norveç) bu oran ortalama yaklaşık %15 civarındadır. Ülkemizde ise bu oran %38,06’dır. Araştırmanın bulguları alan yazındaki pek çok çalışma ile paralellik göstermektedir. Başarıdaki varyansın okullardan kaynaklanan kısmının yüksek olması tek başına yorumlanmamalıdır. Ancak bu farkın yüksek olması, farklı okullarda öğrenim görme durumunun öğrenci başarısını etkilemesi bağlamında incelenerek çalışmalar eğitimde fırsat eşitsizliği kapsamında geliştirilebilir. İleri çalışmalarda başarıdaki varyansa odaklanılarak öğrenci ve okul düzeyinde başka değişkenlerin modellere dâhil edilmesi ve öğrencilerin başarılarındaki varyansın kaynaklarının açıklanması adına çalışmalar yapılması önerilmektedir.

*Anahtar Kelimeler – TIMSS 2019, 8. Sınıf, Fen Başarısı, Hiyerarşik Doğrusal Modelleme, Okullar Arası Fark*

### I. GİRİŞ

Bilimsel ve teknolojik gelişmeler sosyal yaşamın şekillenmesinde büyük rol oynamaktadır. Sosyal yaşamda gereksinim duyulan ve duyulacak bilgi ve beceriler bireylere eğitimin ilk yıllarından itibaren kazandırılmaya çalışılmaktadır. Bilimsel ve teknolojik gelişmelerin etkisiyle hızla artan bilgi, diğer pek çok sistemi etkilediği gibi eğitim sistemlerini de etkilemektedir. Ülkeler, bilimsel ve teknolojik gelişmelere ayak uydurabilmek, gereksinim duyulan bilgi ve becerileri bireylere kazandırabilmek adına eğitim sistemlerinde güncellemeler yapmaktadır.

Bir ülkedeki eğitimin kalitesinin, o ülkenin sahip olduğu eğitim programının niteliği ile doğrudan ilişkisi olduğu söylenebilir. Ülkelerin eğitim programlarında fen eğitimi yer almaktadır ve fen eğitiminde, sürekli değişen ve gelişen fen / bilim çağına ayak uydurabilecek bilimsel donanımı bulunan, teknolojik gelişmelerden her alanda yararlanabilecek ve yaşamın her alanında bilimin gerekliliğinin farkında olan bireyler yetiştirmek, öğretim programının amaçları arasında yer almaktadır.

Olayları araştıran, fikirleri inceleyen, üretken bireyler yetiştirebilmek için fen öğretiminin gerekliliği ve önemi bilinmektedir [1]. Bu nedenle,

öğrencilere küçük yaşlardan itibaren fen becerilerinin kazandırılması gerekmektedir. Becerilerin ölçülmesi konusu ise ölçme ve değerlendirmenin önemli bir konusudur. Öğrencilerin fen bilimlerinde gösterdiği performans, öğretmenlerin sınıf-içi ölçme uygulamalarıyla, ulusal eğitim araştırmalarıyla (ABİDE gibi), ulusal yüksek beklentili sınavlarla (ülkemizdeki kademeler arası geçiş sınavları gibi) ve uluslararası geniş ölçekli testlerle (PISA, TIMSS gibi) sürekli ölçümlenmekte ve izlenmektedir. Okullarda performans kavramı yeni ve özgün bir ürün ortaya koyma ile ilişkilendirildiği [2] gibi, öğrencilerin akademik başarı puanları da performansın bir göstergesi olarak ele alınmaktadır.

Bugün sıklıkla gözlemlendiği üzere, psikolojik yapılar birbiriyle ilişkilidir ve fen başarısının da pek çok değişkenden etkilendiği söylenebilir. Fen eğitiminin bir çıktısı olarak fende başarılı olmanın önemli görülmesinden dolayı ([3], [4], [1]) fen başarısını etkileyen değişkenler sık sık araştırma konusu olarak ele alınmaktadır ([5] - [10]).

Geniş ölçekli testlerden biri olan TIMSS'te ölçümlenen ana alanlardan biri fen başarısıdır. Bu uygulamada fen başarısı (8. sınıf), dört öğrenme alanından (biyoloji, kimya, fizik, yer bilimi) ve üç bilişsel alandan (bilme, uygulama, akıl yürütme) alınan puanlarla ifade edilmektedir. Başarı puanları madde tepki kuramı (item response theory) modelleri ile kestirilmektedir ve her öğrenci için performansı gösteren beş olası değer (plausible value) raporlanmaktadır [11].

Olası değerler, öğrencilerin testteki maddelere verdikleri cevaplara dayalı olarak, öğrencilerin sahip olabileceği yetenek dağılımını ifade etmektedir ve  $\theta$  (theta) değerleri için sonsal olasılık dağılımından rastgele değerlerin çekilmesiyle elde edilmektedir [12]. Rutkowski, Gonzales, Joncas ve von Davier [13] olası değerlerin tek başına kullanılarak ya da bu değerlerin ortalaması alınarak gerçekleştirilen analizlerin yanlı (biased) sonuçlar ürettiğini ifade etmektedir. Laukaityte ve Wiberg [14] de birden fazla olası değer kullanılmasının tahmin doğruluğunu artırdığını ve hatayı azalttığını belirtmektedir. Olası değerler, öğrencilerin bireysel puanları olarak ele alınmamalıdır. Olası değerler, gözlem verilerinden ve ölçüm modellerinden elde edilen kanıtlara dayalı olarak gizil kişi parametrelerinin yerleştirilebilecekleri değer

aralıklarının bir temsilidir [15]. Bu nedenle olası değerlerin birlikte ele alınabildiği modellemeler ile analizlerin gerçekleştirilmesinin daha uygun olduğu düşünülmektedir.

Okulların kendi içerisinde benzeşik (homojen) yapıda olması, fen başarısı açısından okullar arasında başarı farklılıklarının ortaya çıkmasına yol açabilmektedir. Okullar arasındaki bu farklılıkların yakından incelenmesi, öğrenci başarı düzeylerinde gözlenen varyansın kaynaklarının daha iyi anlaşılmasına yardımcı olacaktır. Bu kapsamda okullarda verilen eğitimde, amaçlanan beceriler açısından öğrenciler arası farkların, okullar arası farklar da dikkate alınarak değerlendirilmesi, daha ayrıntılı ve derinlemesine çözüm üretebilme imkânı sağlayabilir.

Bu çalışmada, TIMSS 2019 uygulamasına katılan tüm ülkelerdeki 8. sınıf öğrencilerinin fen başarılarındaki varyansın okullar arası farklılıklardan kaynaklanan kısmının hesaplanması amaçlanmıştır. Dikkate alınan temel araştırma sorusu; “TIMSS 2019 sonuçlarına göre uygulamaya katılan ülkelerdeki okullar arası fark, 8. sınıf öğrencilerinin fen başarı puanlarındaki toplam varyansın ne kadarını açıklamaktadır?” şeklinde belirlenmiştir.

## II. YÖNTEM

İkincil veri analizleri ile yürütülen bu araştırma ilişki arayan araştırma düzeneğinde [16] ilişki araştırma türündedir. İlişkisel araştırmalarda, iki veya daha fazla değişken arasındaki birlikte değişimin varlığının ve derecesinin belirlenmesi amaçlanmaktadır ([17], [18]).

### A. Örneklem

Bu çalışmanın örneklemini TIMSS'in 2019 döngüsüne 8. sınıf düzeyinde katılan ülkelerdeki (39 ülke) 7363 okuldaki 227345 öğrenciden oluşmaktadır. Katılımcı ülkelerde okul sayıları 98 ile 623, öğrenci sayıları 3265 ile 22334 arasında değişmektedir.

### B. Veri Toplama Araçları

Bu çalışmada ikincil veri analizleri, TIMSS'in 2019 yılındaki uygulamasından elde edilen veriler ile gerçekleştirilmiştir. Bu verilere internet üzerinden <https://timss2019.org/international-database/> adresinden erişilebilmektedir. Analizlerde TIMSS'in sunduğu, ilgili değişkenlere ait verileri içeren öğrenci ve okul veri setleri kullanılmıştır.

### C. Veri Analizi

Öğrencilerin okullarda, okulların ülkelerde yuvalanması örneğinde olduğu gibi, hiyerarşik yapıdaki verilerin çözümlenmesinde çok düzeyli modeller kullanılmaktadır. Çok düzeyli istatistiksel modeller için sıklıkla kullanılan HLM genel olarak dört alt model içermektedir. Bu modeller sırasıyla (1) tesadüfi etkili tek yönlü ANOVA modeli, (2) sonuçların ortalamalar olduğu regresyon modeli, (3) tesadüfi katsayılı regresyon modeli ve (4) sabit ve eğim katsayılarının çıktığı modeldir [19]. Bu araştırmada öğrencilerin fen başarı puanlarındaki varyansın okullar tarafından açıklanan kısmının incelenmesi amaçlandığı için tesadüfi etkili tek yönlü ANOVA modeli analizlerde kullanılmıştır. Bu modelde ilk düzeyde öğrencilerin, ikinci düzeyde okulların yer aldığı iki düzey tanımlanmıştır.

Veri analizlerinden önce veri setlerine Arıkan, Özer, Şeker ve Ertaş'ın önerileri [20] dikkate alınarak ağırlık değişkenleri (hem öğrenci, hem okul düzeyinde) eklenmiştir. Birinci düzey (öğrenci) örneklem ağırlıkları TIMSS 2019 veri setlerinde yer alan WGTADJ2 \* WGTFAC2 \* WGTADJ3 \* WGTFAC3 (CLASS WEIGHT ADJUSTMENT \* CLASS WEIGHT FACTOR \* STUDENT WEIGHT ADJUSTMENT \* STUDENT WEIGHT FACTOR) değişkenlerinin çarpımı; ikinci düzey (okul) örneklem ağırlıkları WGTADJ1 \* WGTFAC1 (SCHOOL WEIGHT ADJUSTMENT \* SCHOOL WEIGHT FACTOR) değişkenlerinin çarpımı ile elde edilerek veri setlerine eklenmiştir. 1 ve 2. düzeyde elde edilen ağırlık değişkenlerinin çarpımı tam olarak TIMSS tarafından paylaşılan TOTWGT (TOTAL STUDENT WEIGHT) değerini vermektedir. Değişkenlerin eklenmesi ve modele yönelik artık değerlerin incelenmesi SPSS Statistics 22 [21], diğer veri analizleri HLM 6.04 programı [22] kullanılarak gerçekleştirilmiştir. Modellemenin yapıldığı HLM 6.04 programında .sav uzantılı dosyaların çalıştırılabilmesi için veri setleri .dta (Stata Version 6) uzantısı ile kaydedilerek analizler gerçekleştirilmiştir.

Artık değerlerin (residuals) 0 ortalama ile normal dağılımlı ve homojen olması, düzeylere yönelik hataların hem kendi içlerinde hem de birbirleri arasında ilişkisiz olması temel varsayımları dikkate alınmıştır. Kurulan modellerde elde edilen artıkların dağılımı grafiksel yöntemlerle incelenmiştir ve grafikler, normal dağılımdan ciddi

bir sapma olmadığını göstermektedir. Ayrıca istatistiksel modellemelerin yapılmasının ardından elde edilen artıkların çarpıklık (skewness) değerleri [-0,42; 0,09] aralığında, basıklık (kurtosis) değerleri ise [-0,24; 1,98] aralığında değişmektedir. Buna göre ciddi bir normallik ihlali bulunmadığı yorumu yapılmıştır. Varyansların homojenliğinin testi,  $\chi^2$  istatistiğine dayalı olarak gerçekleştirilmiştir. Bu istatistiğin Tip I hataya duyarlı ve büyük örneklerde istatistiksel olarak anlamlı fark çıkarmaya eğilimli olduğu bilinmektedir. Bu nedenle, artıkların normal dağılım gösteriyor olması yeterli görülmüş ve analizlere devam edilmiştir.

Hiyerarşik doğrusal modellerde örneklem büyüklüğü de önemli bir konudur. Regresyon analizinin genişletilmiş bir hali [23] olan HLM'nin önemli avantajlarından biri gruplandırılmış analiz, araştırmacıyı Tip I hataya karşı koruması iken, önemli dezavantajlarından biri büyük örneklem gerektirmesidir [24]. Özellikle birinci düzeydeki varyansı açıklamak için ikinci düzeydeki grup sayısının yeterli olması gerekmektedir. Bu çalışmada verileri analiz edilen ülkelerdeki okul sayıları 98 ile 623 arasında değişmektedir ve örneklem büyüklüğü HLM için yeterli görülmüştür.

Bu çalışmada kurulan modellerde çıktı değişkeni (bağımlı değişken) olarak öğrencilerin fen başarı puanlarını ifade eden beş olası değer (BSSSCI01-05) birlikte kullanılmıştır.

### III. BULGULAR

Bu araştırmada öğrencilerin 1. düzeyde, okulların 2. düzeyde olduğu, çıktı değişkeni (bağımlı değişkeni) olarak fen başarısının (beş olası değer ile ifade edilmekte) ele alındığı tesadüfi etkili tek yönlü ANOVA modeli (random effects one way ANOVA model) ile analizler yürütülmüştür. Analiz sonucunda her ülke için regresyon denklemleri elde edilmiştir. Bu denklemlerin genel yapısı aşağıdaki gibidir.

$$FenBaşarısı_{ij} = \gamma_{00} + u_{0j} + r_{ij}$$

Yukarıdaki denkleme göre kestirilen katsayılar, bu katsayıların manidarlık testi sonuçları ve fen başarı puanlarındaki varyansın ne kadarlık kısmının okullar arasındaki farklılıklardan kaynaklandığı (yüzde olarak) Tablo 1'de verilmiştir.

**Tablo 1.** Modellere ait istatistikler ve fen başarısındaki varyansın okullar arası farklardan kaynaklanan kısmı

Ülke Kodları	Ülke	Katsayı	Standart Hata	<i>t</i>	sd	$\chi^2$	Okullar Arası Farklardan Kaynaklanan Kısım (%)	Okul Sayısı	Öğrenci Sayısı
USA	ABD	490,786	<b>22,923</b>	21,410	272	<b>8345,63</b>	<b>58,29</b>	273	8698
ARE	BAE	466,003	3,836	121,486	<b>622</b>	<b>31627,73</b>	<b>57,54</b>	<b>623</b>	<b>22334</b>
ZAF	Güney Afrika	353,598	6,446	54,852	<b>518</b>	<b>18450,98</b>	<b>53,42</b>	<b>519</b>	<b>20829</b>
LBN	Lübnan	356,138	7,101	50,156	203	4793,00	50,22	204	4730
HKG	Hong Kong	495,947	6,649	74,589	135	3236,34	49,68	136	3265
ENG	İngiltere	511,944	7,329	69,853	135	2720,65	46,68	136	3365
SGP	Singapur	<b>599,898</b>	5,023	119,433	152	4226,61	46,10	153	4853
IRN	İran	437,222	6,622	66,030	219	4880,42	43,54	220	5980
ISR	İsrail	508,115	6,017	84,452	156	2853,73	41,93	157	3731
MYS	Malezya	464,661	5,260	88,338	176	4120,12	39,99	177	7065
KWT	Küveyt	443,740	6,238	71,133	170	3059,62	39,55	171	4574
EGY	Mısır	391,046	<b>9,291</b>	42,088	168	4185,47	39,02	169	7210
HUN	Macaristan	511,740	5,486	93,280	153	3100,65	38,67	154	4569
ROM	Romanya	455,055	5,313	85,647	197	3046,90	38,10	198	4494
TUR	Türkiye	499,498	7,634	65,432	180	2592,50	38,06	181	4077
QAT	Katar	478,474	5,390	88,767	151	2207,51	37,02	152	3884
MAR	Fas	398,629	5,686	70,106	250	4960,55	36,74	251	8458
JOR	Ürdün	455,005	7,159	63,561	234	4271,49	36,65	235	7176
NZL	Yeni Zelanda	494,740	<b>8,432</b>	58,677	133	2886,95	35,77	134	6051
RUS	Rusya	537,044	4,454	120,565	203	2099,82	30,99	204	3901
KAZ	Kazakistan	480,603	5,417	88,720	167	2323,89	30,95	168	4453
GEO	Gürcistan	448,170	6,086	73,643	144	1414,98	29,79	145	3315
SAU	Suudi Arabistan	427,181	4,868	87,759	208	2286,12	28,76	209	5680
AUS	Avustralya	525,830	5,505	95,514	<b>283</b>	3977,29	27,94	<b>284</b>	<b>9060</b>
CHL	Şili	450,707	3,672	122,757	163	1882,80	27,85	164	4115
BHR	Bahreyn	488,149	5,376	90,797	111	2197,12	27,63	112	5725
LTU	Litvanya	517,861	4,360	118,769	193	1502,08	25,25	194	3826
OMN	Umman	457,185	4,392	104,105	227	2504,48	24,75	228	6751
PRT	Portekiz	516,269	3,934	131,226	155	1192,00	23,98	156	3377
SWE	İsveç	518,685	5,765	89,972	149	1173,00	23,08	150	3996
IRL	İrlanda	517,885	4,866	106,436	148	1177,08	21,16	149	4118
FRA	Fransa	488,140	3,333	146,438	149	1114,46	20,57	150	3874
ITA	İtalya	500,096	2,960	168,924	157	944,19	17,85	158	3619
CYP	Kıbrıs	484,156	3,757	128,870	97	705,49	15,70	98	3521
TWN	Çin Taipei	<b>560,043</b>	3,414	164,053	202	1156,34	15,25	203	4915
FIN	Finlandiya	540,418	2,996	<b>180,365</b>	153	864,97	11,48	154	4874
NOR	Norveç	495,717	4,638	106,881	156	769,21	11,01	157	4575
JPN	Japonya	<b>567,036</b>	2,960	<b>191,560</b>	141	654,75	8,98	142	4446
KOR	Kore	556,040	2,713	<b>204,957</b>	167	462,32	6,30	168	3861

[a] Kurulan tüm modeller ,001 düzeyinde istatistiksel olarak anlamlıdır.

[b] Katsayı, Standart Hata ve *t* sütunlarındaki değerler, robust standart hatalarla birlikte raporlanan son kestirim değerleridir.

[c] Ülkeler, okullar arası farklardan kaynaklanan kısım (%) sütununa göre büyükten küçüğe sıralanmıştır.

[d] Sayısal değer içeren sütunlardaki en büyük üç değer kalın, en küçük üç değer eğik yazı ile belirtilmiştir

Modellemeler sırasında elde edilen tüm güvenilirlik değerleri ,63 (Kore) ile ,98 (ABD ve BAE) arasında değişmektedir. Bu değerler kurulan modellerin ve elde edilen kestirimlerin güvenilirlik düzeyinin yeterli ve çoğunlukla yüksek olduğunu (en düşük üç değer sırası ile ,63; ,77 ve ,80'dir) göstermektedir.

#### IV. TARTIŞMA

Çalışmanın bulgularına göre öğrencilerin fen başarısındaki varyansın okullar arası farklardan kaynaklanan kısmı en yüksek olan ülke ABD ve en düşük olan ülke Kore'dir. Öğrencilerin fen başarısındaki varyansın okullar arası farklılıklardan kaynaklanan kısmı %6,3 ile %58,29 arasında değişmektedir.

Kuzey ülkeleri (Finlandiya, İsveç, Norveç) akademik başarı açısından okullar arası farkların düşük olduğu ülkeler olarak bilinmektedir [25] ve farklı çalışmalarda bu ülkeler için sıklıkla düşük oranlar raporlanmaktadır ([26] - [28]). Bu araştırmada da, kuzey ülkelerindeki öğrencilerin fen başarısındaki varyansın okullar arası farklılıktan kaynaklanan kısmı ortalama yaklaşık %15 olarak hesaplanmıştır.

Avrupa Birliği (AB) üye devletlerindeki öğrencilerin fen başarısındaki varyansın okullar arasındaki farklılıktan kaynaklanan kısmı %11,48 ile %38,67 arasında değişmektedir. En yüksek okullar arası farklılık Macaristan'da iken en düşük okullar arası farklılık Finlandiya'dadır. AB ülkelerinde öğrencilerin fen başarısındaki varyansın okullar arası farklılıktan kaynaklanan kısmı ortalama yaklaşık %23,6'dır.

TIMSS'in 2019'daki döngüsüne (8. sınıf düzeyinde) Türk Cumhuriyeti ülkelerinden yalnızca iki tanesi (Kazakistan ve Türkiye) katılmıştır. Kazakistan'daki öğrencilerin fen başarısındaki varyansın okullar arasındaki farklardan kaynaklanan kısmı yaklaşık %31 iken, Türkiye'de bu oran %38,06'dır. Ülkemiz, genel olarak okullar arası başarı farklılıklarının görece yüksek olduğu ülkelerden biri olarak bilinmektedir ([29] - [31]) ve bu araştırmanın bulguları alan yazındaki bulguları bu açıdan desteklemektedir.

#### V. SONUÇLAR

Araştırmada ulaşılan bulgulara göre, ülkelerin farklı eğitim sistemlerinin bir getirisi olarak öğrencilerin fen başarısındaki varyansın okullar arası farklılıklardan kaynaklanan kısmı çeşitli

derecelerde olabilmektedir. Alan yazında başarı puanlarındaki varyansın okullar arası farklılıktan kaynaklanan kısmının incelenmesi konusuna araştırma sorularında yer veren çeşitli çalışmalar bulunmaktadır ([32], [33], [26], [34], [35], [36], [31], [37]). Öğrenci başarısındaki farklılıkların öğrenci ya da okul özelliklerinden hangileri tarafından açıklanabileceği önemli bir çalışma konusudur. Farklı okullarda öğrenim gören öğrencilerin akademik başarılarının farklılaşıyor olması ve bu farklılaşmanın derecesinin yüksek olması eğitimde fırsat eşitsizliği kapsamında ele alınabilir [38]. İleri çalışmalarda, hem öğrenci düzeyindeki hem okul düzeyindeki varyansların kaynaklarına odaklanılması, bu kapsamda öğrenci ve okul düzeyinde başka değişkenlerin modellemelere dâhil edilmesi önerilmektedir.

#### KAYNAKLAR

- [1] Köseoğlu, F. ve Kavak, N. (2001). Fen öğretiminde yapılandırmacı yaklaşım. *Gazi Üniversitesi Eğitim Fakültesi Dergisi*, 21(1), 139-148.
- [2] Popham, J. W. (2005). *Modern education measurement*. Allyn and Bacon.
- [3] Hançer, A. H., Şensoy, Ö. ve Yıldırım, H. İ. (2003). İlköğretimde çağdaş fen bilgisi öğretiminin önemi ve nasıl olması gerektiği üzerine bir değerlendirme. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 13(13), 80-88.
- [4] Holbrook, J., & Rannikmae, M. (2009). The meaning of scientific literacy. *International Journal of Environmental & Science Education*, 4(3), 275-288. <https://eric.ed.gov/?id=ej884397>
- [5] Allen, L. R. (1973). An evaluation of children's performance on certain cognitive, affective, and motivational aspects of the systems and subsystems unit of the Science Curriculum Improvement Study Elementary Science Program. *Journal of Research in Science Teaching*, 10(2), 125-134. <https://eric.ed.gov/?id=EJ079473>
- [6] Atar, H. Y. (2014). Öğretmen niteliklerinin TIMSS 2011 fen başarısına çok düzeyli etkileri. *Eğitim ve Bilim*, 39(172), 121-137.
- [7] Atar, H. Y. ve Atar, B. (2012). Examining the effects of Turkish education reform on students' TIMSS 2007 science achievements. *Educational Sciences: Theory & Practice*, 12(4), 2632-2636. <https://files.eric.ed.gov/fulltext/EJ1002867.pdf>
- [8] Becker, B., & Chang, L. (1986). Measurement of science achievement and its role in gender differences. Paper presented at the Annual Meeting of the American Educational Research Association (70th, San Francisco, CA, April 16-20, 1986). <https://files.eric.ed.gov/fulltext/ED270328.pdf>
- [9] Hough, L. W., & Piper, M. K. (1982). The relationship between attitudes toward science and science achievement. *Journal of Research in Science Teaching*, 19(1), 33-38. <https://doi.org/10.1002/tea.3660190105>

- [10] Zhang, F., Bae, C. L., & Broda, M. (2021). Science self-concept, relatedness, and teaching quality: A multilevel approach to examining factors that predict science achievement. *International Journal of Science and Mathematics Education*, 20(3), 503-529. <https://doi.org/10.1007/s10763-021-10165-2>
- [11] Mullis, I. V. S., & Martin, M. O. (Eds.). (2017). *TIMSS 2019 Assessment Frameworks*. Retrieved from Boston College, TIMSS & PIRLS International Study Center. <http://timssandpirls.bc.edu/timss2019/frameworks/>
- [12] Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation*, 31(1-2), 114-128. <https://doi.org/10.1016/j.stueduc.2005.05.005>
- [13] Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International large-scale assessment data: Issues in secondary analysis and reporting. *Educational Researcher*, 39(2), 142-151. <https://doi.org/10.3102/0013189X10363170>
- [14] Laukaiyte, I., & Wiberg, M. (2017). Using plausible values in secondary analysis in large-scale assessments. *Communications in Statistics – Theory and Methods*, 46(22), 11341-11357. <https://doi.org/10.1080/03610926.2016.1267764>
- [15] Mislevy, R. J. (1993). Should “multiple imputations” be treated as “multiple indicators”? *Psychometrika*, 58(1), 79-85. <https://doi.org/10.1007/BF02294472>
- [16] Erkuş, A. (2009). *Davranış bilimleri için bilimsel araştırma süreci* (2. Baskı). Seçkin Yayıncılık.
- [17] Creswell, J. W. (2012). *Educational research: Planning, conducting and evaluating quantitative and qualitative research* (4th Ed.). Pearson.
- [18] Fraenkel, J. R., & Wallen, N. E. (2006). *How to design and evaluate research in education* (6th Ed.). McGraw-Hill.
- [19] Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods 2*. Sage Publications.
- [20] Arıkan, S., Özer, F., Şeker, V. ve Ertaş, G. (2020). Geniş ölçekli testlerde örneklem ağırlıklarının ve olası değerlerin önemi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 11(1), 43-60. <https://doi.org/10.21031/epod.602765>
- [21] IBM Corp. (2013). IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.
- [22] Raudenbush, S., Bryk, A., & Congdon, R. (2007). HLM for Windows (Version 6.04). Lincolnwood, IL: Scientific Software International.
- [23] Huta, V. (2014) When to use hierarchical linear modeling. *The Quantitative Methods for Psychology*, 10(1), 13-28. <https://doi.org/10.20982/tqmp.10.1.p013>
- [24] McCoach, D. B. (2010). Hierarchical linear modeling. In G. R. Hancock, & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (123-140). Routledge.
- [25] Välijärvi, J., & Malin, A. (2003). *The two-level effect of socio-economic background*. In P. Linnakylä, S. Lie, & A. Roe (Eds.), Northern lights on PISA (pp. 123-131). University of Oslo.
- [26] Güngör M. ve Bekmezci, S. M. (2022). Matematik başarısı açısından okullar arası farklar: TIMSS 2019. *1st International Conference on Scientific and Academic Research (ICSAR)* içinde (377-381. ss.). Konya, Türkiye.
- [27] Mohammadpour, E., & Abdul Ghafar, M. N. (2014). Mathematics achievement as a function of within- and between-school differences. *Scandinavian Journal of Educational Research*, 58(2), 189-221. doi:10.1080/00313831.2012.725097
- [28] OECD. (2016). *PISA 2015 Results (Volume I): Excellence and Equity in Education*. OECD Publishing.
- [29] Aydın, M. (2015). *Öğrenci ve okul kaynaklı faktörlerin TIMSS matematik başarısına etkisi* [Doktora Tezi, Necmettin Erbakan Üniversitesi]. Konya, Türkiye.
- [30] Suna, H. E., Özer, M., Şensoy, S., Gür, B. S., Gelbal S. ve Aşkar, P. (2021). Türkiye’de akademik başarının belirleyicileri. *Journal of Economy Culture and Society*, 64, 143-162. <https://doi.org/10.26650/JECS2021-934211>
- [31] Yavuz, H. Ç., Demirtaşlı, R. N., Yalçın, S. ve İlgün Dibek, M. (2017). Türk öğrencilerin TIMSS 2007 ve 2011 matematik başarısında öğrenci ve öğretmen özelliklerinin etkileri. *Eğitim ve Bilim*, 42(189), 27-47. <http://dx.doi.org/10.15390/EB.2017.6885>
- [32] Demirkol, S. ve Kelecioğlu, H. (2022). İngilizce başarısına etki eden öğrenci ve okul özelliklerinin hiyerarşik lineer model ile incelenmesi. *Eğitim ve Yeni Yaklaşımlar Dergisi*, 5(1), 96-109. <https://doi.org/10.52974/jena.1126043>
- [33] Demirkol Karakuş, S. (2017). Öğrenci, öğretmen ve okul özelliklerinin TEOG matematik başarıları üzerindeki etkilerinin hiyerarşik doğrusal modellerle incelenmesi [Yüksek Lisans Tezi, Hacettepe Üniversitesi]. Ankara, Türkiye.
- [34] Güngör, M. ve Demir, E. (2022). Farklı ülkelerde 8. Sınıf öğrencilerinin matematik başarısı açısından okullar arası farkların incelenmesi. *Uluslararası Eğitim Kongresi 2022 (International Education Congress, EduCongress)* içinde (345-350. ss.). Akdeniz Üniversitesi, Antalya, Türkiye. <https://educongress.org/wp-content/uploads/2022/12/TM.2022.pdf>
- [35] Güngör, M., Gören, S. ve Doğan, N. (2022). Türk Cumhuriyetlerinde TIMSS 2019 matematik başarısı ve okullar arası başarı farkları. *8. Uluslararası Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi (CMEEP) 2022* içinde (321-326. ss.). Ege Üniversitesi, İzmir, Türkiye.
- [36] Suna, H. E. ve Özer, M. (2021). Türkiye’de sosyoekonomik düzey ve okullar arası başarı farklarının akademik başarı ile ilişkisi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 12(1), 54-70. <https://doi.org/10.21031/epod.860431>
- [37] Yavuz, S., Odabaş, M. ve Özdemir, A. (2016). Öğrencilerin sosyoekonomik düzeylerinin TEOG matematik başarısına etkisi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 7(1), 85-95. <https://doi.org/10.21031/epod.86531>
- [38] Berberoğlu, G. ve Kalender, İ. (2005). Öğrenci başarısının yıllara, okul türlerine, bölgelere göre incelenmesi: ÖSS ve PISA analizi. *Eğitim Bilimleri ve Uygulama*, 4(7), 21-35.