

Makul Değer ve Ağırlıklandırma Kullanımının Doğrusal Regresyon ve Hiyerarşik Doğrusal Model Parametrelerine Etkisi*

Osman TAT **

İlhan KOYUNCU ***

Selahattin GELBAL ****

Öz

Uluslararası Öğrenci Değerlendirme Programı (PISA) ve Uluslararası Matematik ve Fen Eğilimleri Çalışması (TIMSS) gibi geniş ölçekli uygulamalarda öğrenci yeteneğine ilişkin kestirimler olarak makul değerler kullanılır. Bu çalışmalarda katılımcılar tabakalı örnekleme yöntemi ile çekilmektedir. Bu durum elde edilen verilerin, çok sayıda tabakadan oluşan hiyerarşik bir yapıda olmasının önünü açmaktadır. Geniş ölçekli değerlendirme çalışmaları bağlamında makul değerler, sonsal yetenek dağılımından rastgele elde edilen değerler olarak tanımlanmaktadır. Doğrusal modellerde tek bir makul değer veya tüm makul değerlerin ortalamasının bağımsız değişken olarak kullanılmasının yanlı sonuçlara sebep olabileceği bilinmektedir. Aynı zamanda bu geniş ölçekli çalışmaların verileri ile analizler yapılırken örnekleme ağırlıklarının göz ardı edildiği sıkça gözlenmektedir. Bu çalışmanın amacı, çoklu doğrusal regresyon ve hiyerarşik doğrusal modellerde 1) tek makul değer kullanımının, 2) tüm makul değerlerin kullanımının, 3) ağırlık kullanma durumunun parametre kestirimlerine etkisini araştırmaktır. Çalışmada PISA 2015 uygulamasının Türkiye verilerinden yararlanılmıştır. Araştırmada, örnekleme ağırlıkları kullanma durumunun ve makul değerlerin kullanım şeklinin kat sayıların, standart hataların ve açıklanan varyans oranının kestirilmesinde önemli rolleri olduğu belirlenmiştir. Bulgular detaylı bir biçimde tartışılmış ve uygulama ve gelecek araştırmalar için bazı öneriler sunulmuştur.

Anahtar Kelimeler: Hiyerarşik doğrusal modeller, çoklu doğrusal regresyon, makul değer, örnekleme ağırlıkları, geniş ölçekli değerlendirmeler, PISA.

GİRİŞ

Birçok ülkede, eğitsel girişimler ve politikalar geliştirmek amacıyla grup performansını belirlemede geniş ölçekli uygulamalardan elde edilen veriden yararlanılır. Bu uygulamalarda, öğrenci performansını ölçen bilişsel testlerin yanında, öğrenci, öğretmen ve okul düzeyinde bilgi toplanmasını sağlayan ölçekler de kullanılır. Elde edilen veriler öğrencilerin okul ve çalışma durumları ile ilgili becerileri ile bir arada ele alınarak bireysel değerlendirme yerine grup düzeyinde çıkarımlarda bulunulur. Bu tür geniş çaplı uygulamalarda çok geniş konu alanlarına ilişkin performansın ölçülmesinden kaynaklı zaman kaybını önlemek amacıyla farklı kitapçıklar geliştirilerek ikili bloklar halinde öğrencilere uygulanır. Bu durumda, öğrencilerin tümü aynı sorulara cevap vermedikleri için klasik istatistiksel yöntemlerle performanslarının kestirilip grup düzeyinde karşılaştırılmaları yanlı ve hatalı olacaktır (Organization for Economic Cooperation and Development-OECD, 2017). Bu nedenle, bu tarz uygulamalarda öğrenci yeterliliğinin olası dağılımını gösteren çoklu değerler kullanılmaktadır (Von Davier, Gonzales & Mislevy, 2009). *Makul değer (Plausible value)* olarak adlandırılan bu

* Bu çalışmanın bulguları 6. Uluslararası Eğitim Konferansında sözlü bildiri olarak sunulmuştur, Zagreb, Hırvatistan, 2017.

** Arş. Gör., Hacettepe Üniversitesi, Eğitim Fakültesi, Ankara-Türkiye, osmntt@gmail.com, ORCID ID: 0000-0003-2950-9647

*** Dr. Öğr. Üyesi, Adıyaman Üniversitesi, Eğitim Fakültesi, Adıyaman-Türkiye, ilhankync@gmail.com, ORCID ID: 0000-0002-0009-5279

**** Prof. Dr., Hacettepe Üniversitesi, Eğitim Fakültesi, Ankara-Türkiye, sgelbal@gmail.com, ORCID-ID: 0000-0001-5181-7262

Bu makaleye atıfta bulunmak için:

Tat, O., Koyuncu, İ., & Gelbal, S. (2019). The influence of using plausible values and survey weights on multiple regression and hierarchical linear model parameters. *Journal of Measurement and Evaluation in Education and Psychology*, 10(3), 235-248. doi: 10.21031/epod.486999

Geliş Tarihi: 23.11.2018

Kabul Tarihi: 17.06.2019

değerler, öğrencilerin aldıkları alt testlere verdikleri cevapların yanında, duyuşsal özelliklere ve ulaşılabilir geçmiş bilgilerine (demografik bilgiler) de dayanır (Mislevy, 1991; OECD, 2009).

Makul değerler, geniş ölçekli uygulamalar bağlamında, yetenek puanının sonsal dağılımlardan yapılan rastgele seçimler anlamına gelmektedir (Von Davier ve diğerleri., 2009). Madde Tepki Kuramı modellerinden olan Rasch modeli (Rasch, 1960) ile yapılan kestirimlerde kullanılan En Yüksek Olabilirlik (EO), Ağırlıklandırılmış En Yüksek Olabilirlik (AEO) (Warm, 1985), Birleşik En Yüksek Olabilirlik (BEO) (Wright & Stone, 1979) ve Beklenen Önsel (BÖ) (Bock & Aitkin, 1981) kestirim yöntemleri birbirlerinin eksik yönlerini kapatan tekniklerdir. Ancak bu yöntemler nokta kestirimler yapmakta ve makul değerlerde olduğu gibi bireyler için sonsal dağılımından gelen birbirinden farklı birden fazla yetenek kestirimi vermemektedir (Wu, 2005). Makul değer kullanımından ilk kez 1994 yılında yapılan U.S. National Assessment of Educational Progress (NAEP) uygulamasından elde edilen verilerin analizinde yararlanılmıştır. Bu kullanım şekli Rubin (1987) tarafından yapılan çoklu yerleştirme (multiple imputation) çalışmasından esinlenilmiştir. Sonraki yıllarda yapılan NAEP uygulamaları ve OECD tarafından yapılan Trends in International Mathematics and Science Study (TIMSS) ve the Programme for International Student Assessment (PISA) çalışmalarında kullanılmasıyla birlikte geniş ölçekli sınavlarda makul değer kullanımı daha yaygın hale gelmiştir. Bu uygulamalarda genel olarak, her öğrenci için beş makul değer üretilir ancak literatürde bu sınırlama ile ilgili güçlü dayanak bir bulunmamaktadır (Von Davier ve diğerleri., 2009; Wu, 2005).

Makul değerler, öğrencilerin maddelere verdikleri cevaplara dayanarak bir öğrencinin sahip olabileceği yetenek dağılımına karşılık gelir ve Madde Tepki Kuramında (MTK) yer alan θ yetenek değerleri için sonsal olasılık dağılımından rastgele değerlerin çekilmesiyle elde edilir (Wu, 2005). 1983-1984 yıllarında yapılan NAEP ve 2000 yılında yapılan PISA uygulaması teknik raporlarında bu değerlerin nasıl hesaplandığı ve olasılık dağılımından nasıl çekildiği ile ilgili ayrıntılı bilgiler yer almaktadır (Adams & Wu, 2002; Beaton, 1987). Mislevy (1993), makul değerlerin öğrencilerin geleneksel anlamda kullanılan bireysel puanı olmamalarından dolayı aynı puanın ya da gizil değişkenin çoklu göstergeleri gibi incelenmemesi gerektiğine vurgu yapmaktadır. Von Davier ve diğerleri (2009) yaptıkları çalışmada, makul değer kullanımının grup düzeyinde yapılacak değerlendirmelerde nokta kestirimi yapan EAP ve WMLE yöntemlerine göre daha düşük hata miktarına sahip sonuçlar verdiğini göstermişlerdir. Ancak makul değerlerin ortalamasının (PV-W) kullanımının, bu değerlerin her biri ile yapılan analizlerin sonucunda elde edilen istatistiklerin ortalamasının (PV-R) kullanımına göre daha yanlış kestirimler yaptığı ve kullanılmaması gerektiğini ifade etmişlerdir (Von Davier ve diğerleri., 2009). Ayrıca, Wu (2005) tarafından yapılan simülasyon çalışmasında makul değerlerden herhangi birinin tek başına kullanılmasının evren parametrelerinin yüksek oranda doğrulukla kestiriminde yeterli olduğu görülmüştür.

TIMSS, PISA, NAEP ve ICILS (International Computer and Information Literacy Study) gibi geniş ölçekli çalışmalarda her öğrenci için yetenek puanı olarak nokta kestirimi yerine öğrenci yeteneğinin sonsal dağılımından elde edilen makul değerler kullanılmaktadır. Geniş ölçekli bu uygulamalarda elde edilen veri birden çok düzeyin (öğrenci, okul, bölge, ülke, vb.) bulunduğu hiyerarşik bir yapıdadır. Aslında, sosyal bilimlerde alanında organizasyon araştırmaları, kültürlerarası çalışmalar, gelişimsel araştırmalar gibi birçok araştırma türünde bu veri yapısına rastlamak mümkündür (Bryk & Raudenbush, 2002). Eğitim bilimlerinde de veriler öğrenciler veya herhangi bir analiz birimi için elde edilen tekrarlı ölçümlerin yanı sıra öğrencilerin sınıf, sınıfın okul, okulun şehir ya da bölge içinde yuvalanması (nested) şeklinde iki ve daha fazla düzeyde olabilir. Örneğin, Chowa, Masa, Ramos, ve Ansong (2015), Ghana Youth Save verisi üzerinde öğrencilere ve okullara ait özelliklerin, gençlerin akademik başarısını nasıl etkilediğini incelemişlerdir. Bu çalışmada, öğrenciler okullar içinde yuvalanmıştır. Stipek ve Valentino (2015), National Longitudinal Survey of Youth (NLSY)'e katılan öğrencilere ait boylamsal verileri kullanarak, okul öncesi öğrencilerin kısa süreli ve işleyen bellek ile dikkat değişkenlerine ait ölçümlerin matematik ve okuduğunu anlamadaki boylamsal gelişimlerini nasıl yordadığını incelemişlerdir. Zaman içinde yapılan ölçümler, ikinci düzey birim olan öğrenci değişkeninde yuvalanmış durumdadır. Bryk ve Raudenbush (1988) ise The Sustaining Effects Study (SES) çalışmasında birinci sınıftan üçüncü sınıfa kadar katılan öğrencilerin yoksulluk ve okul

yoksulluk yoğunluğunun okuduğunu anlama ve matematikteki öğrenmeleri ile olan ilişkisini incelemek amacıyla üç düzeyli bir hiyerarşik doğrusal model kullanılmışlardır.

Verilerin hiyerarşik yapısı göz önünde bulundurulmadığında verilerin iki farklı şekilde kullanıldığına sıklıkla rastlanır. Bu kullanım şekilleri birleştirme (aggregation) ve ayrıştırma (disaggregation). Birleştirme yönteminde, verinin alt birimleri daha üst birimlerde birleştirilir. Öğrencilerin bir testten aldıkları puanların sınıf düzeyinde birleştirilmesi, sınıf düzeyindeki ortalama puanların ağırlıklandırılarak okul düzeyinde puanların elde edilmesi buna örnek verilebilir. Başka bir ifade ile öğrencilerin ortalamasının devam ettiği okullara atanması bir birleştirme örneğidir. Bu yöntemde bireysel farklılıklar göz ardı edildiğinden birleştirme işlemi yapılmış değişkenler arasındaki ilişkiler çok daha güçlü olabilir ve yanlış yorumlamalara yol açabilir (Bryk & Raudenbush, 2002; Snijders & Bosker, 2003; Woltman, Feldstain, MacKay, & Rocchi, 2012). Ayrıştırma yönteminde ise üst düzey birimler daha alt düzeylere ayrıştırılır. Örneğin, okul ve sınıf düzeyinde bir değişkene ait verilerin öğrencilere atanması buna örnek verilebilir. Bu durumda, aynı okulda veya sınıfta yer alan öğrencilerin tümü aynı özellikte olacağından istatistiksel analizlerin önemli bir varsayımı olan gözlemlerin bağımsızlığı ilkesi ihlal edilmiş olacaktır (Snijders & Bosker, 2003; Woltman ve diğerleri, 2012). Sonuç olarak, birleştirme ve ayrıştırma işlemleri ile doğrusal regresyon modelleri kullanıldığında, değişkenlere ait hataların ilişkili olmasına, grup içi farklılıkların göz ardı edilerek regresyon denklemindeki katsayıların ve standart hataların hatalı kestirilmesine neden olacaktır (Bryk & Raudenbush, 2002).

Yuvalanmış veriyi incelemenin yolu ise birleştirme ve ayrıştırma yöntemlerinin bu dezavantajlarını elimine eden hiyerarşik doğrusal modellerdir. Geçmişte önemli bir sorun olan analiz birimini ve ölçüm değişimini incelemenin önündeki engeller hiyerarşik doğrusal modellerin kullanılması ile birlikte ortadan kalkmıştır (Bryk & Raudenbush, 1988). Bu sayede, her bir düzeydeki değişkenlere ait kestirimler, aynı ve farklı düzeydeki değişkenler arasındaki etkileşimler ve varyans-kovaryans bileşenleri tek bir analiz yardımı ile incelenebilmektedir (Bryk & Raudenbush, 2002). Hiyerarşik veride hiyerarşik doğrusal modellerin kullanılmasının düzey içi ve düzeyler arasındaki ilişkilerin doğru bir şekilde formüle edilmesi, birleştirmeden kaynaklı yanlılıkların elimine edilmesi, ampirik araştırmalarda daha çeşitli ve kapsamlı araştırma sorusu ve hipotez ileri sürme imkanı tanınması, rastgele etkileri içeren uygun hata yapılarının belirlenmesi ve grup etkilerinden kaynaklı varyans ve kovaryans bileşenlerini içeren standart hata kestirimine imkan sağlaması gibi birçok avantajı vardır (Raudenbush, 1988). Goldstein'e (2011) göre, hiyerarşik modellerin regresyon katsayılarının istatistiksel olarak kullanışlı kestirilmesini, doğru standart hatalar, güven aralıkları ve anlamlılık testleri elde edilmesini, düzey içi ve düzeyler arası ilişkilerin incelenebilmesini ve tüm faktörlerin göz önünde bulundurulması düzeylerin karşılaştırılabilmesini sağlar. Hiyerarşik Lineer Modeller (HLM) analizlerinin nasıl yapıldığı ve istatistiksel boyutuna bu araştırmanın veri analizi kısmında değinilmiştir.

Verilerin hiyerarşik yapısının göz ardı edilmesi bazen sonuçların önemli ölçüde farklılaşmasına neden olmaktadır. Roberts (2004) çalışmasında, verinin hiyerarşik yapısı göz önünde bulundurulmadığı durumda şehirlik (urbanicity) ile öğrencilerin fen başarısı arasındaki ilişkinin .77 olduğunu tespit etmişken, öğrencilerin okullarda yuvalandığı göz önünde bulundurulduğunda bu ilişkinin -.88 olduğunu belirtmiştir. Benzer bir biçimde birçok çalışmada da hiyerarşik modeller yerine geleneksel doğrusal modeller kullanmanın hatalı sonuçlar vereceği ifade edilmiştir (Bryk & Raudenbush, 2002; Goldstein, 2011; Osborne, 2000; Raudenbush, 1988; Raudenbush & Bryk, 1986; Woltman ve diğerleri., 2012). Çoklu doğrusal regresyon ve hiyerarşik doğrusal modellerin karşılaştırıldığı çalışmada Atar (2010), ikinci düzey olan okullar arasında *Fen'e Karşı Tutum* değişkenine ait katsayısının -0.2 ile 1.09 arasında değerler aldığını belirtmiştir. Bu bulgular okullar arasında Fen'e karşı tutum düzeyinin belirgin bir biçimde farklılaştığını ve verilerin çok düzeyli yapısının göz önünde bulundurulmasının önemini ortaya koymuştur.

Gelman (2006), klasik regresyon analizi ile karşılaştırıldığında düzeyi değişmekle birlikte hiyerarşik modellerin hemen her zaman, yordamada olmazsa olmaz, veri indirgemede yararlı ve nedensel çıkarımlarda yardımcı bir gelişme olduğunu vurgulamıştır. Buna karşın hiyerarşik veride, hiyerarşik

doğrusal modellerin kullanımı parametre kestiriminin yansızlığını garantilememektedir. Çünkü diğer doğrusal modellerde olduğu gibi bazı durumlarda seçilen örneklemin uygulama evrenindeki öğrenci sayısını temsil etmemesinden kaynaklı kestirim hataları gözlenebilir. Bu amaçla, geniş ölçekli sınavlarda farklı düzeylere ait ağırlıklandırmalardan yararlanılır (Meinck, 2015).

PISA gibi geniş ölçekli sınavlarda kullanılan ağırlıklandırmalar, örneklem çekme hatalarının doğru hesaplanmasını, evrene yönelik doğru kestirimlerin ve yorumların yapılmasını sağlayarak verinin analizini kolaylaştırılmaktadır. Böylece, kullanıcıların, katılımcı her bir ülkenin karmaşık örneklem yapısı için yanlı olmayan standart hata kestirimlerinin ve anlamlılık testlerinin yapılarak uygun güven aralıklarının oluşturulması sağlanmaktadır. Ağırlıklandırma işleminde, seçilen her okulun tam temsilini sağlamak, okul mevcudiyetinin belirli oranlardaki katılımını dengelemek, bazı okullardaki düşük cevaplama oranlarını göz önünde bulundurmamak, küçük grupların büyük ağırlıklandırma değerlerini önlemek ve bazı ülkelerde uygulanan anketlere yönelik seçilen fazladan öğrenci sayısının etkisi dengelemek amaçlandığından aynı ülkedeki öğrencilerin ağırlık değerleri farklılık göstermektedir (OECD, 2014). TIMSS, PIRLS ve PISA gibi sınavlarda anket ağırlıklandırmaları için temel alınan istatistiksel prosedürler, Cochran (1977), Lohr (2010) ile Sarndal, Swensson ve Wretman'ın (1992) çalışmalarında görmek mümkündür. Bu sınavlarda yapılan ağırlıklandırma işlemleri, okul temelli ve okul-İçi temelli ağırlıklandırma faktörleri ile 5 tane denkleştirme (adjustment) faktörü üzerinden yapılmaktadır. Denkleştirme faktörleri, belirli bir okuldaki katılımcılara benzer özelliklere sahip diğer okullardaki katılımcı olmayan öğrencileri göz önünde bulundurmamak, öğrencilerin yaş ve sınıf düzeylerini dengelemek, okul İçi katılımcı olmayan öğrencileri cinsiyet, sınıf düzeyi ve bölge durumuna göre göz önünde bulundurmamak, beklenmedik çok yüksek okul temelli ve diğer tüm ağırlıklandırma faktörlerini düşürmek amacıyla kullanılır. PISA 2012 uygulaması için bu ağırlıklandırma faktörlerinin nasıl hesaplandığı ile ilgili detaylı bilgilere teknik rapordan (OECD, 2014) ulaşılabilir.

Çalışmanın Amacı

Geniş ölçekli sınavlarda kullanılan makul değerler ve ağırlıklandırma değerleri yapılan ölçümlerin daha hassas ve kapsayıcı olmasını temel alır. Bu bağlamda, bu çalışmanın amacı 2015 yılında yapılan PISA uygulamasının öğrenci verisi kullanılarak, öğrencilerin fen bilgisi okuryazarlığının yordamasında çoklu doğrusal regresyon ve hiyerarşik doğrusal model analizlerinin sonuçlarını makul değerler ve ağırlıklandırma kullanma durumlarına göre karşılaştırmaktır. Bu genel amaç çerçevesinde öncelikle ağırlıklandırma olmaksızın, bir makul değer bağımlı değişken olduğu çoklu doğrusal regresyon ve HLM analizleri gerçekleştirilmiştir. Ardından, tüm makul değerlerin bağımlı değişken olduğu ve ağırlıklandırmaların yapıldığı çoklu doğrusal regresyon ve HLM analizleri gerçekleştirilmiştir. Bu iki yaklaşım yardımı ile hem makul değer kullanım şeklinin ve ağırlıklandırmaların etkisi hem de çoklu doğrusal regresyon ve HLM analizi sonuçları karşılaştırılabilmektedir. Buna göre a) ağırlıklandırmanın olmadığı ve bir makul değer bağımlı değişken olduğu durumda, b) ağırlıklandırmanın yapıldığı ve tüm makul değerlerin bağımlı değişken olduğu durumda çoklu doğrusal regresyon ve HLM analizi sonuçları;

1. Herhangi bir bağımsız değişken yokken nasıldır?
2. Birinci düzey bağımsız değişkenler olarak öğrencilerin fen bilimleri epistemolojik inançları, test kaygıları, motivasyonları ve ekonomik, sosyal ve kültürel statü indeksi modellendiğinde nasıldır?
3. İkinci düzey bağımsız değişkenler olarak okullardaki sınıf mevcudu, eğitime yönelik liderlik, eğitim materyali yetersizliği ve eğitim personeli yetersizliği modellendiğinde nasıldır?
4. Birinci ve ikinci düzey bağımsız değişkenler birlikte modellendiğinde nasıldır?

YÖNTEM

Bu araştırma, geniş çaplı uluslararası eğitim araştırmalardan elde edilen hiyerarşik yapıdaki verilerde, makul değerler, ağırlıklar ve bazı bağımsız değişkenler arasındaki ilişkiyi iki farklı analiz yönteminde gösterilmeyi amaçlayan, ilişkisel tarama modelinde (Fraenkel, Wallen & Hyun, 2012) nicel bir çalışmadır.

Çalışma Grubu

Çalışmanın amacı doğrultusunda PISA'ya ait 2015 yılı verilerinden yararlanılmıştır. PISA 15 yaş grubundaki öğrencilerin temel olarak matematik, fen bilgisi ve okuduğunu anlama alanlarındaki performanslarını ölçmek için OECD tarafından her üç yılda bir yürütülen uluslararası bir uygulamadır. İlki 1997 yılında gerçekleştirilen çalışmanın sonuncusu 2015 yılında yapılmış ve 72 ülkeden 520 bine yakın öğrenciden veri toplanmıştır. PISA'ya Türkiye'den 187 okuldan toplam 5895 öğrenci katılmıştır.

Bu çalışmada hiyerarşik doğrusal modellerin doğası gereği olarak iki düzeyli hiyerarşik veri (birinci düzey öğrenciler ve ikinci düzey okul) kullanılmıştır. İkinci düzey örnekleme Türkiye örnekleminde eksik verisi bulunmayan 178 okuldan oluşmaktadır. HLM'de ikinci düzey birimlerde eksik veriler ile analiz yapılması mümkün olmadığından eksik verisi bulunan okullar veri setinden çıkartılmıştır. HLM 7 yazılımı ikinci düzey verilerin eksiksiz olmasını zorunlu kılmaktadır. Eksik veri varlığında eksik gözlemler için ya bir değer atanmalıdır veya bu gözlemler silinmelidir. Eksik veri durumu göz ardı edilip analizler yapılmaya çalışıldığında yazılım sistem dosyalarını oluştururken eksik verisi bulunan tüm gözlemleri otomatik olarak silmektedir (Palardy, 2011). Araştırmanın birinci düzey örnekleme ise belirtilen 178 okulda eğitim gören toplam 5703 öğrenciden oluşmaktadır. Maas ve Hox (2005) hiyerarşik doğrusal modellerin sağlıklı sonuçlar verebilmesi için en az 50 ikinci düzey biriminin gerekli olduğunu vurgulamışlardır. Bu bakımdan araştırmanın çalışma grubu hiyerarşik doğrusal modellere ilişkin analizlerin yürütülmesi için uygun büyüklüktedir.

Veri Toplama Araçları

PISA'da öğrencilere matematik, fen ve okuduğunu anlama testleri uygulanarak öğrencilerin bu alanlardaki bilişsel beceri düzeyleri incelenmektedir. Bu bilişsel beceri testlerinin yanında, her uygulamada bu üç alandan biri odak alan olarak belirlenir ve odak alana ilişkin olarak duyuşsal değişkenlerin incelendiği öğrenci anketi uygulanır. Yani öğrencilere ilişkin veriler bilişsel testler ve duyuşsal değişkenlerin araştırıldığı anketler aracılığı ile toplanır. Benzer bir biçimde PISA'da okulların teknik altyapı yeterliği, eğitimsel kaynaklarının durumu gibi çok çeşitli konularda bilgi elde etmek için okul yöneticilerine okul anketi uygulanır. Bu çalışmada öğrenci anketi ve fen bilgisi testinden elde edilen birinci değişkenler ile okul anketinden elde edilen ikinci düzey değişkenler beraber kullanılmıştır. Çalışmada kullanılan değişkenler Otomatik Doğrusal Modelleme (Yang, 2013) yardımı ile belirlenmiştir. 14 değişkenle yürütülen bu analizde önce en önemli on değişken tespit edilmiş, ardından bu on değişkenden sekiz tanesinin (ortak önem derecesine sahip iki değişkenden birer tanesi kullanılmıştır) çalışmada kullanılması kararlaştırılmıştır. Kısıtlı sayıdaki değişkenle oluşturulan modeller yardımı ile bulguların daha açık bir biçimde ortaya konacağı düşünülmüştür. Tablo 1'de belirlenen bu değişkenlere ilişkin detaylar görülmektedir.

Tablo 1. Değişkenler ve Özellikleri

Seviye	Değişken	Kısaltma	Türü	Yapısı
Öğrenci (Birinci Seviye)	Fen Bilgisi Okuryazarlık Puanı (Makul Değerler, 10 adet)	PV1SCIE (1-10)	Bağımlı	Sürekli
	Epistemolojik İnanç	EPIST	Bağımsız	Sürekli
	Test Kaygısı	ANXTEST	Bağımsız	Sürekli
	Başarım Motivasyonu	MOTIVAT	Bağımsız	Sürekli
	Ekonomik, Sosyal ve Kültürel Düzey İndeksi	ESCS	Bağımsız	Sürekli
Okul (İkinci Seviye)	Okulun Ortalama Sınıf Mevcudu	CLSIZE*	Bağımsız	Sürekli
	Öğretmen Katılımı	LEADTCH*	Bağımsız	Sürekli
	Okulun Eğitim Materyali Eksiği	EDUSHORT*	Bağımsız	Sürekli
	Okulun Eğitim Personeli Eksiği	STAFFSHO*	Bağımsız	Sürekli
Ağırlıklar	Düzenlenmiş Öğrenci Düzeyi Ağırlık	W_FSTUWT		Sürekli
	BRR-FAY Ağırlık Replikasyonları (80 adet)	W_FSTURWT (1-80)		Sürekli

*Çoklu regresyon analizinde öğrenci düzeyine ayrıştırılmıştır (disaggregation)

Verilerin Analizi

Makul değerler ve ağırlıkların farklı istatistiksel analizlerdeki etkisini inceleyebilmek için bu çalışmanın analizinde çoklu regresyon ve hiyerarşik doğrusal modeller analizi kullanılmıştır. Çoklu doğrusal regresyon analizinde öncelikle bağımlı değişken PV1SCIE1 olarak belirlenmiş ve hiçbir açıklayıcı değişkenin olmadığı, sadece öğrenci düzeyi değişkenlerin olduğu, sadece okul düzeyi değişkenlerin olduğu ve her iki düzeye ait değişkenin beraber modellendiği dört farklı model analiz edilmiştir. Çoklu regresyon analizinde okul değişkenleri öğrencilere atanmıştır (disaggregation). Ardından, bahsi geçen bu dört model, 10 makul değer (PV1SCIE1-10) bağımlı değişken olduğu, öğrenci düzeyi ağırlığın ve replikasyonların (W_FSTUWT ve W_FSTURWT1-80) kullanıldığı durumda tekrar edilmiştir. Böylelikle makul değerler ve ağırlık kullanımının çoklu regresyon analizine etkisi incelenmiştir.

HLM analizlerinde de çoklu regresyon analizinde olduğu gibi sadece birinci makul değer (PV1SCIE1) bağımlı değişken olduğu dört model ve tüm makul değerlerin ve ağırlığın kullanıldığı dört model olmak üzere toplam sekiz model test edilmiştir. Verilerin analizinde çoklu regresyonda ağırlıkların ve tüm makul değerlerin kullanımını sağlayan betiğin (syntax) hazırlanmasında IDB Analyzer (International Association for the Evaluation of Educational Achievement-IEA, 2016) programı kullanılmıştır. Asıl analizler ise SPSS 21.0 (International Business Machines-IBM Corp., 2012) ve HLM 7 Hierarchical Linear and Nonlinear Modelling (Bryk, Raudenbush & Congdon, 2010) programları yardımı ile gerçekleştirilmiştir. Analizlerin tamamında .05 manidarlık düzeyi olarak kabul edilmiştir.

Çoklu regresyon ve HLM analizleri yapılmadan önce her iki analiz için de varsayımların karşılanıp karşılanmadığı incelenmiştir. İlk olarak çoklu regresyon analizinin varsayımları olan bağımlı değişken ve bağımsız değişkenler arasında doğrusal bir ilişki olması, bağımsız değişkenler arasında çoklu bağlantı sorunu olmaması, artık değerlerin bağımsızlığı (ilişkisiz olması), artık varyanslarının sabit olması (homoscedasticity), artık değerlerin normal dağılması ve aykırı değer bulunmaması bağımsız değişkenlerin kullanıldığı tüm modellerde (ağırlık kullanılan ve kullanılmayan ilk modeller tamamen koşulsuz modellerdir) test edilmiştir. Buna göre tüm modellerde de bağımlı değişkene karşı bağımsız değişkenlerin olduğu saçılım grafikleri ve bu saçılıma en iyi uyan doğru çizdirilmiş ve doğrusal bir ilişki tanımlanabildiği tespit edilmiştir. Çoklu bağlantı durumu tolerans değeri ve VIF değerleri yardımı ile incelenmiştir. Tüm modellerde tolerans değerlerinin 0.2'den büyük olduğu (0.7 - 0.9) ve VIF değerlerinin 10'dan küçük olduğu (1.3 - 1.5) belirlenmiştir. Artık değerlerin bağımsızlığı için Durbin-Watson istatistiği incelenmiş birden küçük ve üçten büyük bir değere rastlanmamıştır. Artık değerlerin sabit olduğu (homoscedastic), kestirilen standart puanlar ile standart artık değerlerden oluşturulan grafik yardımı ile tespit edilmiştir. Oluşturulan P-P grafiğinde artık değerlerin köşegen üzerinde doğrusal bir biçimde dağıldığı görülmüştür. Bu durumda artık değerlerin normal dağıldığı söylenebilir. Son olarak tüm çoklu regresyon modelleri için aykırı değerlerin varlığı Cook'un uzaklığı yöntemi ile incelenmiş ve birden büyük herhangi bir değere saptanmamıştır.

HLM analizlerinin birinci modeli tamamen koşulsuz model (null) olduğundan varsayımlar bakımından incelenmemiştir. Diğer tüm modellerin ise hata varyanslarının homojenliği ve hataların normal dağılımı yönünden test edilmesi gerekmektedir (Snijders ve Bosker, 1999). Buna göre tüm modellerde birinci düzey hata terimlerinin ikinci düzeyler arasındaki dağılımı saçılım grafikleri yardımı ile incelenmiş ve varyansın ikinci düzeyler arasında rastgele dağıldığı yani homojen olduğu belirlenmiştir. Aynı zamanda yine tüm modellerin birinci ve ikinci düzey standart artık değerleri ile kestirilen standart değerler P-P grafikleri ile incelenmiş ve her iki düzeydeki artık değerlerin normal dağıldığı belirlenmiştir.

BULGULAR

Birinci Alt Probleme İlişkin Bulgular

Herhangi bir açıklayıcı değişkenini olmadığı durumda kurulan dört farklı modele ilişkin ayrıntılar Tablo 2’de görülmektedir. Tablo 2 incelendiğinde Türkiye genel fen okuryazarlığı ortalamasının en yüksek kestirildiği model tüm makul değerlerin ve ağırlıkların kullanıldığı çoklu regresyon modeli iken, en düşük kestirildiği model tüm makul değerlerin ve ağırlıkların kullanıldığı HLM analizidir. En küçük standart hata kestirimi çoklu regresyon modeli (1.02) ile elde edilirken, en yüksek standart hata, tüm makul değerlerin ve ağırlıkların kullanıldığı HLM analizi ile elde edilmiştir.

Tablo 2. Birinci Modele İlişkin Sabit Etkiler

Analiz	Sabit Etki	Katsayılar	SH	t
Çoklu Regresyon (PV1SCIE1)	Genel Fen Okuryazarlığı Ortalaması	423.19*	1.02	414.89
Çoklu Regresyon (PV1SCIE1-10) Ağırlıklandırılmış	Genel Fen Okuryazarlığı Ortalaması	426.22*	4.06	104.98
HLM (PV1SCIE1)	Genel Fen Okuryazarlığı Ortalaması, γ_{00}	418.48*	4.35	96.13
HLM (PV1SCIE1-10) Ağırlıklandırılmış	Genel Fen Okuryazarlığı Ortalaması, γ_{00}	417.71*	4.90	85.29

*p < .05

İki farklı rastgele etkiler ANOVA modelinden elde edilen rastgele etkiler Tablo 3’te sunulmuştur. Her iki analiz sonucuna göre öğrencilerin fen okuryazarlığı ortalamaları okullar arasında farklılaşmaktadır. Ayrıca her iki analizde kestirilen birinci düzey hata terimi çok yakın iken, tüm makul değerlerin ve ağırlıkların kullanıldığı HLM analizinde kestirilen ikinci seviye hata terimi ilk analize göre daha yüksektir.

Tablo 3. Birinci Modele İlişkin Rastgele Etkiler

Analiz	Rastgele Etki	ss	Varyans	χ^2
HLM (PV1SCIE1)	2.Seviye Hata Terimi, u_{0j}	55.43	3073.53	5499.68*
	1.Seviye Hata Terimi, r_{ij}	53.61	2873.90	
HLM (PV1SCIE1-10) Ağırlıklandırılmış	2.Seviye Hata Terimi, u_{0j}	59.46	3536.05	6332.42*
	1.Seviye Hata Terimi, r_{ij}	53.37	2848.21	

*p < .05

Fen okuryazarlığındaki varyansın okul seviyesinde açıklanabilecek oranını belirlemek için sınıf-içi korelasyon katsayısından yararlanılmıştır. Buna göre her iki analizden elde edilen oranlar şu şekildedir:

$$\rho_1 = \frac{\tau_{00}}{(\tau_{00} + \sigma^2)} = 3073.53 / (3073.53 + 2873.90) = 0.517 \quad (1)$$

$$\rho_2 = \frac{\tau_{00}}{(\tau_{00} + \sigma^2)} = 3536.05 / (3536.05 + 2848.21) = 0.554 \quad (2)$$

Birinci analizde bağımlı değişkendeki varyansın yaklaşık olarak %52'sinin ($\rho_1 = 0.517$) okul düzeyi ile açıklanabildiği tespit edilmiştir. Buna karşın tüm makul değerlerin kullanıldığı durumda bağımlı değişkendeki varyansın yaklaşık %55'inin ($\rho_2 = 0.554$) ikinci düzey ile açıklanabildiği belirlenmiştir.

İkinci Alt Probleme İlişkin Bulgular

Tablo 4'te öğrenci düzeyine ait dört değişkenin modele dahil edildiği iki farklı çoklu regresyon analizine ait katsayılar ve iki farklı rastgele katsayı modelinden elde edilen sabit etkiler sunulmuştur.

Tablo 4. İkinci Modele İlişkin Sabit Etkiler

Analiz	Sabit Etki	Katsayılar	SH	t
Çoklu Regresyon (PV1SCIE1)	Genel Fen Okuryazarlığı Ortalaması	452.97*	1.65	274.11
	EPIST	14.74*	0.83	17.80
	ANXTEST	-6.81*	0.94	-7.29
	MOTIVAT	6.05*	0.98	6.16
	ESCS	17.84*	0.82	21.66
Çoklu Regresyon (PV1SCIE1-10) Ağırlıklandırılmış	Genel Fen Okuryazarlığı Ortalaması	456.05*	4.56	100.08
	EPIST	15.23*	1.3	11.75
	ANXTEST	-6.27*	1.4	-4.47
	MOTIVAT	6.53*	1.38	4.74
	ESCS	18.69*	2.05	9.12
HLM (PV1SCIE1)	Genel Fen Okuryazarlığı Ortalaması, γ_{00}	423.31*	4.61	91.74
	EPIST, γ_{10}	7.37*	0.68	10.81
	ANXTEST, γ_{20}	-6.35*	0.82	-7.72
	MOTIVAT, γ_{30}	3.26*	0.90	3.63
	ESCS, γ_{40}	1.91*	0.78	2.46
HLM (PV1SCIE1-10) Ağırlıklandırılmış	Genel Fen Okuryazarlığı Ortalaması, γ_{00}	422.10*	5.18	81.49
	EPIST, γ_{10}	7.41*	0.85	8.71
	ANXTEST, γ_{20}	-6.15*	0.93	-6.61
	MOTIVAT, γ_{30}	3.89*	1.05	3.70
	ESCS, γ_{40}	1.87*	0.91	2.03

*p < .05

Tablo 4 incelendiğinde tüm analizlerde genel fen okuryazarlığının yükseldiği söylenebilir. Bu analizlerdeki katsayılar bağımsız değişkenler kontrol altında tutulduğunda fen okuryazarlığı ortalamasını yansıtmaktadır. Tüm analizlerde, tüm bağımsız değişkenler bağımlı değişkeni manidar bir biçimde yordamaktadır. Çoklu regresyon analizlerinde en yüksek katsayıya sahip değişken ekonomik, sosyal ve kültürel düzey indeksiyken (ESCS), bu değişken HLM analizlerinde oldukça küçük kestirilmiştir. HLM analizlerinde öğrencilerin epistemolojik inanç düzeyleri (EPIST) en yüksek katsayıya sahip değişkendir. Ayrıca, çoklu regresyon analizlerinde tüm katsayıların HLM analizlerine göre büyük kestirildiği söylenebilir. Ağırlıklandırılmamış çoklu regresyon ve HLM analizlerinde katsayılar ait standart hataların ağırlıklandırılmış çoklu doğrusal regresyona göre daha küçük kestirildiği belirlenmiştir. Tablo 5'te rastgele katsayı modellerinden elde edilen rastgele etkiler görülmektedir.

Tablo 5. İkinci Modele İlişkin Rastgele Etkiler

Analiz	Rastgele Etki	ss	Varyans	χ^2
HLM (PV1SCIE1)	2.Seviye Hata Terimi, u_{0j}	56.50	3192.56	1982.77*
	EPIST Etkisi, u_{1j}	1.51	2.28	143.22
	ANXTEST Etkisi, u_{2j}	3.94	15.51	159.71
	MOTIVAT Etkisi, u_{3j}	5.26	27.69	176.79
	ESCS Etkisi, u_{4j}	1.68	2.82	146.86
	1.Seviye Hata Terimi, r_{ij}	52.15	2719.55	
HLM (PV1SCIE1-10) Ağırlıklandırılmış	2.Seviye Hata Terimi, u_{0j}	60.19	3622.63	2220.32*
	EPIST Etkisi, u_{1j}	2.47	6.09	144.92
	ANXTEST Etkisi, u_{2j}	2.72	7.38	146.64
	MOTIVAT Etkisi, u_{3j}	5.16	26.59	163.78
	ESCS Etkisi, u_{4j}	3.02	9.11	149.17
	1.Seviye Hata Terimi, r_{ij}	51.89	2693.07	

*p < .05

Rastgele etkilerde modele birinci düzey bağımsız değişkenlerin eklenmesi ile birinci düzey hata varyanslarının küçülmesi beklenir. Modele dahil edilen birinci düzey değişkenlerin birinci düzeye ait varyansı hangi oranda açıkladıklarını belirlemek için Eşitlik 3 ve Eşitlik 4'ten yararlanılmıştır.

$$\text{Ağırlıklanılmamış HLM } \rho_1 = (\sigma^2_{ANOVA} - \sigma^2_{RIM}) / \sigma^2_{ANOVA} = (2873.90 - 2719.55) / 2873.90 = 0.05(3)$$

$$\text{Ağırlıklandırılmış HLM } \rho_2 = (\sigma^2_{ANOVA} - \sigma^2_{RIM}) / \sigma^2_{ANOVA} = (2848.21 - 2693.07) / 2848.21 = 0.05 \quad (4)$$

Her ne kadar tüm makul değerlerin ve ağırlıkların kullanıldığı HLM analizinde birinci düzey hataya ait varyans daha küçük olsa da kurulan her iki HLM modeli birinci düzey varyansı eşit oranda açıklamaktadır. Ağırlıkların kullanıldığı durumda ikinci düzeye ilişkin hata varyansı daha yüksek kestirilmiştir.

Üçüncü Alt Probleme İlişkin Bulgular

Tablo 6'da okul düzeyine ait dört değişkenin öğrenci düzeyine ayrıştırıldığı (disaggregation) iki farklı çoklu regresyon analizine ait katsayılar ve farklı iki HLM analizine ilişkin sabit etkiler görülmektedir.

Tablo 6. Üçüncü Modele İlişkin Sabit Etkiler

Analiz	Sabit Etki	Katsayılar	SH	t	p
Çoklu Regresyon (PV1SCIE1)	Genel Fen Başarı Ortalaması	411.89	4.621	89.14	.00
	CLSIZE	0.290	0.09	3.07	.00
	LEADTCH	5.00	0.90	5.56	.00
	EDUSHORT	-9.58	0.94	-10.25	.00
	STAFFSHO	-8.43	1.01	-8.35	.00
Çoklu Regresyon (PV1SCIE1-10) Ağırlıklandırılmış	Genel Fen Okuryazarlığı Ortalaması	416.84	25.41	16.4	.00
	CLSIZE	0.27	0.53	0.51	.00
	LEADTCH	4.38	5.2	0.84	.00
	EDUSHORT	-10.43	3.75	-2.78	.00
	STAFFSHO	-11.05	4.3	-2.57	.00
HLM (PV1SCIE1)	Genel Fen Okuryazarlığı Ortalaması, γ_{00}	399.33	16.37	24.40	.00
	CLSIZE, γ_{10}	0.52	0.35	1.50	.14
	LEADTCH, γ_{20}	4.05	3.65	1.11	.27
	EDUSHORT, γ_{30}	-9.09	3.02	-3.01	.00
	STAFFSHO, γ_{40}	-9.85	4.01	-2.45	.02
HLM (PV1SCIE1-10) Ağırlıklandırılmış	Genel Fen Okuryazarlığı Ortalaması, γ_{00}	399.89	20.49	19.51	.00
	CLSIZE, γ_{10}	0.57	0.42	1.35	.18
	LEADTCH, γ_{20}	3.38	4.42	0.77	.45
	EDUSHORT, γ_{30}	-8.44	3.70	-2.29	.02
	STAFFSHO, γ_{40}	-15.39	4.63	-3.32	.00

*p < .05

Tablo 6 incelendiğinde çoklu regresyon analizlerinin her ikisinde de tüm değişkenlerin bağımlı değişkeni manidar bir biçimde yordadığı, buna karşın HLM analizlerinde ise sadece eğitim kaynakları yetersizliğinin (EDUSHORT) ve eğitim personeli yetersizliğinin (STAFFSHO) manidar olduğu söylenebilir. Hem regresyon analizlerinde hem de HLM analizlerinde ağırlıklandırılmış makul değerlerin kullanımı ile katsayılara ilişkin standart hataların arttığı söylenebilir. Ağırlıklandırılmış çoklu regresyon katsayılarının bazılarının ağırlıklandırılmamış çoklu regresyon analizi katsayılarından az da olsa büyük olduğu görülmektedir. HLM analizlerinde ağırlıklandırmanın katsayıları fazla etkilemediği belirlenmiştir. Tablo 7’de HLM analizlerine ilişkin rastgele etkiler sunulmuştur.

Tablo 7. Üçüncü Modele İlişkin Rastgele Etkiler

Analiz	Rastgele Etki	ss	Varyans	χ^2
HLM (PV1SCIE1)	2.Seviye Hata Terimi, u_{0j}	51.64	2667.09	4700.68*
	1.Seviye Hata Terimi, r_{ij}	53.61	2873.60	
HLM (PV1SCIE1-10) Ağırlıklandırılmış	2.Seviye Hata Terimi, u_{0j}	53.46	2857.64	5243.57*
	1.Seviye Hata Terimi, r_{ij}	53.36	2847.75	

Tablo 7 incelendiğinde ağırlıklandırılmış HLM analizinde ikinci seviye hata varyansının daha yüksek kestirildiği görülmektedir. Tamamen koşulsuz modele ikinci düzey değişkenlerin eklenmesi ile ikinci düzey varyansın düşmesi beklenir. Modele eklenen ikinci düzey değişkenlerin ikinci düzey varyansı hangi oranda açıkladığını belirlemek için Eşitlik 5 ve Eşitlik 6’dan yararlanılmıştır.

$$\text{Ağırlıklandırılmamış HLM } \rho_1 = (\sigma^2_{ANOVA} - \sigma^2_{MAOR}) / \sigma^2_{ANOVA} = (3073.53 - 2667.09) / 3073.53 = 0.13 \quad (5)$$

$$\text{Ağırlıklandırılmış HLM: } \rho_2 = (\sigma^2_{ANOVA} - \sigma^2_{MAOR}) / \sigma^2_{ANOVA} = (3536.05 - 2847.75) / 3536.05 = 0.20 \quad (6)$$

Dört ikinci düzey bağımsız değişkenin modele eklenmesi sonucunda ağırlıklandırılmamış HLM analizinde ikinci düzey varyansın %13’ü açıklanabilirken, bu oran ağırlıklandırılmış HLM analizinde %19,5’e çıkmaktadır. Ağırlıklandırma işleminin HLM analizinde açıklanan varyans oranını etkilediği söylenebilir.

Dördüncü Alt Probleme İlişkin Bulgular

Tablo 8’de birinci düzey değişkenlerle manidar olduğu belirlenen ikinci düzey değişkenlerin birlikte modellendiği iki farklı regresyon modeline ait katsayılar ve iki farklı bağımlı değişken olarak kesişim ve eğimler modeline ilişkin sabit etkiler sunulmuştur.

Tablo 8’e göre çoklu regresyon analizlerinde en büyük katsayıya sahip değişkenler öğrenci düzeyi değişkenler olan öğrencilerin epistemolojik inanç düzeyleri (EPIST) ve ekonomik, sosyal ve kültürel düzey indeksiyken (ESCS), HLM analizlerinde katsayısı en yüksek olan yordayıcılar yine öğrenci düzeyine ait epistemolojik inanç düzeyi (EPIST) ve öğrencilerin test kaygısıdır (ANXTEST). Ayrıca Tablo 8’de daha önceki modellerde olduğu gibi regresyon analizlerindeki katsayıların HLM analizlerine göre daha yüksek kestirildiği, katsayılarla ilişkin standart hataların ise daha düşük kestirildiği görülmektedir. Ayrıca çoklu regresyon analizlerinde, ağırlıklandırma yapma durumunda ikinci düzey değişkenlerin standart hatalarında ciddi bir büyüme gözlenmektedir. HLM analizlerinde ise ağırlıklandırma işlemi katsayılar ve standart hatalarında çok büyük değişim yaratmamıştır. Tablo 9’da HLM analizlerine ilişkin rastgele etkiler görülmektedir.

Tablo 8. Dördüncü Modele İlişkin Sabit Etkiler

Analiz	Sabit Etki	Katsayılar	SH	t
Çoklu Regresyon (PV1SCIE1)	Genel Fen Okuryazarlığı Ortalaması	453.91*	1.65	275.59
	EDUSHORT	-7.11*	0.884	-8.04
	STAFFSHO	-7.00*	0.96	-7.26
	EPIST	14.00*	0.81	17.23
	ANXTEST	-6.79*	0.92	-7.42
	MOTIVAT	5.76*	0.96	5.98
	ESCS	15.04*	0.83	18.17
Çoklu Regresyon (PV1SCIE1-10) Ağırlıklandırılmış	Genel Fen Okuryazarlığı Ortalaması	457.20*	4.79	95.35
	EDUSHORT	-7.8*	3.07	-2.54
	STAFFSHO	-9.07*	3.83	-2.37
	EPIST	14.16*	1.22	11.58
	ANXTEST	-6.32*	1.3	-4.86
	MOTIVAT	6.21*	1.35	4.61
	ESCS	15.63*	1.94	8.07
HLM (PV1SCIE1)	Genel Fen Okuryazarlığı Ortalaması, γ_{00}	429.33*	4.71	91.20
	EDUSHORT, γ_{01}	-8.85*	3.25	-2.72
	STAFFSHO, γ_{02}	-6.97*	3.55	-1.96
	EPIST, γ_{10}	7.43*	0.68	10.85
	ANXTEST, γ_{20}	-6.32*	0.82	-7.69
	MOTIVAT, γ_{30}	3.22*	0.89	3.61
	ESCS, γ_{40}	1.85*	0.78	2.39
HLM (PV1SCIE1-10) Ağırlıklandırılmış	Genel Fen Okuryazarlığı Ortalaması, γ_{00}	431.37*	5.26	81.97
	EDUSHORT, γ_{01}	-8.24*	3.54	-2.33
	STAFFSHO, γ_{02}	-12.35*	3.95	-3.13
	EPIST, γ_{10}	7.43*	0.83	8.92
	ANXTEST, γ_{20}	-6.09*	0.93	-6.52
	MOTIVAT, γ_{30}	3.78*	1.05	3.59
	ESCS, γ_{40}	1.86	0.93	2.00

Tablo 9. Dördüncü Modele İlişkin Rastgele Etkiler

Analiz	Rastgele Etki	ss	Varyans	χ^2
HLM (PV1SCIE1)	2.Seviye Hata Terimi, u_{0j}	52.97	2805.70	1666.33*
	EPIST Etkisi, u_{1j}	1.63	2.66	143.29
	ANXTEST Etkisi, u_{2j}	3.95	15.59	159.68
	MOTIVAT Etkisi, u_{3j}	5.13	26.35	176.69
	ESCS Etkisi, u_{4j}	1.65	2.72	146.78
	1.Seviye Hata Terimi, r_{ij}	52.15	2719.60	
HLM (PV1SCIE1-10) Ağırlıklandırılmış	2.Seviye Hata Terimi, u_{0j}	55.15	3041.55	1786.47*
	EPIST Etkisi, u_{1j}	2.50	6.24	144.95
	ANXTEST Etkisi, u_{2j}	2.70	7.27	146.54
	MOTIVAT Etkisi, u_{3j}	5.10	26.99	163.38
	ESCS Etkisi, u_{4j}	2.96	8.79	149.15
	1.Seviye Hata Terimi, r_{ij}	51.89	2692.70	

Bağımlı değişken olarak kesişim ve eğimler modellerinde açıklanan birinci ve ikinci düzey varyans oranlarının belirlenmesi için bu modellerden elde edilen varyanslar rastgele etkiler ANOVA modelinden elde edilen bulgularla karşılaştırılmıştır.

Açıklanan birinci düzey varyans:

$$\text{Ağırlıklandırılmamış: } \rho_1 = (\sigma^2_{ANOVA} - \sigma^2_{MAOANCOVA}) / \sigma^2_{ANOVA} = (2873.90 - 2719.60) / 2873.90 = 0.05 \quad (7)$$

$$\text{Ağırlıklandırılmış: } \rho_2 = (\sigma^2_{ANOVA} - \sigma^2_{MAOANCOVA}) / \sigma^2_{ANOVA} = (2848.21 - 2692.70) / 2848.21 = 0.05 \quad (8)$$

Açıklanan ikinci düzey varyans:

$$\text{Ağırlıklandırılmamış: } \rho_1 = (\sigma^2_{ANOVA} - \sigma^2_{MAOANCOVA}) / \sigma^2_{ANOVA} = (3073.53 - 2805.70) / 3073.53 = 0.09 \quad (9)$$

$$\text{Ağırlıklandırılmış: } \rho_2 = (\sigma^2_{ANOVA} - \sigma^2_{MAOANCOVA}) / \sigma^2_{ANOVA} = (3536.05 - 3041.55) / 3536.05 = 0.14 \quad (10)$$

Buna göre bağımlı değişken olarak kesişim ve eğimler modeli ile yapılan analizlerde tek makul değer kullanıldığı ve ağırlıklandırmanın yapılmadığı durumlarda açıklanan birinci düzey varyans aynıdır. Buna karşın tüm makul değerlerin ve ağırlıkların birlikte kullanıldığı durumda açıklanan ikinci düzey varyans daha yüksek (%14) bulunmuştur.

SONUÇLAR ve TARTIŞMA

Bu çalışmada makul değerlerin ağırlıklar ile birlikte kullanılması durumunda çoklu doğrusal regresyon ve HLM analizi sonuçlarının karşılaştırılması amaçlanmıştır. Bu amaç doğrultusunda, herhangi bir açıklayıcı değişken olmadığında, öğrenci düzeyi değişkenler olduğunda, okul düzeyi değişkenler olduğunda ve her iki düzeydeki açıklayıcı değişkenler bir arada olduğunda söz konusu yöntemlerin kestirimleri karşılaştırılmıştır.

Herhangi bir açıklayıcı değişkenin yer almadığı modellerde, en yüksek genel fen okuryazarlığı ortalaması tüm makul değerlerin ve ağırlıkların kullanıldığı çoklu doğrusal regresyon modeli ile kestirilmiştir. Genel anlamda her iki çoklu doğrusal regresyon analizinin fen okuryazarlığını HLM analizine göre yüksek kestirdiği söylenebilir. Ağırlıklandırma işleminin ise hem regresyon analizinde hem de HLM analizinde katsayılara ait standart hatanın kestirilmesinde etkili olduğu belirlenmiştir. Buna göre ağırlıklandırma durumunda her iki analizde de standart hataların büyüdüğü gözlenmiştir. Bu durumda ağırlıklandırmaların katsayıların manidarlığı noktasında önemli bir rolü olduğu savunulabilir. Her iki rastgele etkiler ANOVA modelinin rastgele etkileri incelendiğinde öğrencilerin fen okuryazarlığının okullar arasında farklılaştığı belirlenmiştir. Ağırlık kullanma durumunda ikinci düzey birimler olan okullar tarafından açıklanan varyans oranının yükseldiği gözlenmiştir. Bu durum okullar arasındaki farklılığın ağırlıklandırma sonucunda daha da arttığı anlamına gelir. Bu çalışmada bağımlı değişkendeki varyansın yaklaşık olarak %55'ine kadarının ikinci düzey birimler tarafından açıklanabildiği tespit edilmiştir. Bu bulgu birçok çalışmada (Bryk & Raudenbush, 2002; Goldstein, 2011; Osborne, 2000; Raudenbush, 1988; Raudenbush & Bryk, 1986; Woltman ve diğerleri., 2012) vurgulanan HLM modellerinin kullanılmasının önemini bir kez daha göstermiştir.

Tüm modeller için, öğrencilerin fen bilimleri epistemolojik inançları, test kaygıları, motivasyonları ve ekonomik, sosyal ve kültürel düzey indeksi değişkenleri fen okuryazarlığını yordamaktadır. Literatürde PISA veya TIMSS gibi geniş ölçekli sıvalarda bu değişkenlerden özellikle ekonomik, sosyal ve kültürel düzey indeksinin birçok çalışmada incelendiği (Acar ve Öğretmen, 2012; Atar ve Atar, 2012) ve manidar olduğu bilinmektedir. Birinci düzey değişkenlerin incelendiği bu dört modelin üçünde epistemolojik inanç ve ekonomik, sosyal ve kültürel düzey indeksi en büyük katkılı değişkenler olmasına karşın, bu değişkenlere ilişkin katsayılar çoklu regresyon analizlerinde HLM modellerine göre oldukça yüksek kestirilmiştir. Her dört modelde kestirilen standart hatalar birbirine oldukça yakındır. Yalnız ağırlıklandırma durumunda kestirilen standart hataların diğer modellere göre biraz daha yüksek olduğu gözlenmiştir. Buna göre, her ne kadar tüm açıklayıcı değişkenler manidar ve standart hatalar yakın (ağırlıklandırılmamış çoklu regresyon hariç) değerlerde olsa da HLM ve çoklu doğrusal regresyon analizinden elde edilen katsayıların önemli farklılıklar gösterdiği sonucuna ulaşılmıştır. Roberts (2004) tarafından yapılan çalışmada, verinin hiyerarşik yapısı göz önünde bulundurulmadığında araştırma bulgularının önemli düzeyde farklılaştığının gözlemlenmesi bu sonuçla paralellik göstermektedir. HLM analizinde, sadece bir makul değer kullanıldığı model ve tüm makul değerlerin ağırlıklandırmalar ile birlikte kullanıldığı modelin açıkladığı birinci düzey varyans miktarları eşit çıkmıştır. Bu sonuç açıklayıcı değişkenlerin öğrenci düzeyinde, ağırlıkların ise okul düzeyinde olması ile ilişkili olabilir. Ayrıca bu durum, Wu (2005) tarafından yapılan benzetim çalışmasında makul değerlerden herhangi birinin tek başına kullanılmasının evren parametrelerinin yüksek oranda doğrulukla kestiriminde yeterli olduğu sonucu ile de örtüşmektedir.

Okullardaki sınıf mevcudu, eğitime yönelik liderlik, eğitim materyali yetersizliği ve eğitim personeli yetersizliği ikinci düzey açıklayıcı değişkenleri her iki çoklu doğrusal regresyon modelinde anlamlı çıkmıştır. Ancak, her iki HLM modeli için sadece eğitim materyali yetersizliği ve eğitim personeli yetersizliği değişkenleri önemli bulunmuştur. Bu durum, çoklu doğrusal regresyon analizinde düzey farkı göz ardı edildiğinden t değerleri olması gerektiğinden yüksek çıkmakta ve dolayısıyla yuvalanmış veride ikinci düzey değişkenlerin anlamlı çıkma eğilimi göstermesinden kaynaklanmaktadır. Yapılan çalışmalar ve incelemelerde de geleneksel analizlere göre HLM analizinin yordamada daha etkili olduğu ve katsayılar ile bu katsayılarla ilişkin standart hataların daha doğru kestirdiğini ortaya koymuştur (Gelman, 2006; Goldstein, 2011; Raudenbush, 1988). Hem çoklu doğrusal regresyon hem de HLM yönteminde tüm makul değerlerin ağırlıklandırmalarla birlikte kullanılması katsayılarla ilişkin standart hataları artırmıştır. Bu durumda her iki analiz yöntemi için de ağırlıklandırma kullanımının ikinci tip hata riskini düşürdüğü söylenebilir. Tüm makul değerlerin ağırlıklandırmalarla birlikte kullanılması HLM analizinde katsayıları fazla etkilememekle birlikte açıklanan varyans oranını artırmıştır. Buna göre, birden fazla makul değer ve ağırlıklandırmaların kullanılmasının HLM analizinin performansını arttırdığı söylenebilir.

Öğrenci ve okul düzeyinde modele dahil edilen tüm değişkenlerin tüm modellerde öğrencilerin genel fen performanslarını manidar bir biçimde yordayan faktörler olduğu görülmüştür. Bununla birlikte çoklu doğrusal regresyon analizinde en yüksek katsayıya sahip değişkenler öğrencilerin epistemolojik inanç düzeyleri ile ekonomik, sosyal ve kültürel düzey indeksiyken, HLM analizlerinde epistemolojik inanç düzeyi ve okulun eğitim materyali eksiği ve okulun eğitim personeli eksiği değişkenleridir. Ayrıca, tüm değişkenlerin modele dahil edildiği bu durumda bile çoklu doğrusal regresyon analizlerinde HLM analizlerinde göre katsayılar daha yüksek kestirilirken, katsayılarla ilişkin standart hatalar daha düşük kestirilmiştir. Yapılan çalışmalarda da hiyerarşik veride geleneksel doğrusal modeller ile HLM sonuçlarının farklılaştığını görülmüştür (Bryk and Raudenbush, 2002; Gelman, 2006; Goldstein, 2011; Osborne, 2000; Raudenbush, 1988; Raudenbush and Bryk, 1986; Woltman ve diğerleri, 2012). Tüm katsayıların ağırlıklandırmalarla birlikte kullanılması çoklu doğrusal regresyonda okul düzeyi değişkenlerin katsayılarının standart hatalarını önemli düzeyde etkilerken HLM analizlerinde ise önemli bir değişim yaratmamıştır. HLM analizlerinde tüm makul değerlerin ağırlıklandırmalar ile birlikte kullanılması öğrenci ve okul düzeyinde açıklanan varyans oranlarında önceki modellerdekine benzer bir etki yaratmıştır. Buna göre açıklanan öğrenci düzeyi varyansı değişmemiş ancak okul düzeyi açıklanan varyans oranı artmıştır. Buna göre, okul düzeyi değişkenlerinin modele dahil edilmesi sonuçlarda farklı bir etki yaratmaktadır. Bu sonuçlar, düzey farklılaşmasının analizlerde göz önünde bulundurulmasının gerekliliğinin desteklemektedir.

Çoklu doğrusal regresyon analizinde tüm makul değerler ağırlıklandırmalar ile birlikte kullanıldığında, öğrenci ve okul düzeyindeki değişkenler ayrı veya birlikte modellere dahil edildiği durumlarda model katsayılarında önemli bir artış görülmezken, katsayılarla ilişkin standart hatalarda artış görülmektedir. t değerleri standart hataların yükselmesinden etkilendiğinden yordayıcı değişkenlerin anlamlılık düzeyleri bu durumdan etkilenebilmektedir. Bu çalışmada, modele dahil edilen değişkenler t değerlerindeki düşüşe rağmen anlamlı çıkmıştır. Bu nedenle tüm makul değerlerin ağırlıklandırmalar ile birlikte kullanılmasının çoklu doğrusal regresyon parametrelerinde önemli bir değişim yaratmadığı söylenebilir. Bu sonuç makul değer kullanımıyla ilgili Wu (2005) tarafından yapılan çalışmanın sonuçlarıyla paralellik gösterse de ağırlıklandırmaların kullanılmasıyla ilgili OECD (2014) tarafından önerilen kullanım şeklinin sonuçlarda bir değişiklik yaratmadığını göstermiştir. HLM analizinde de katsayılar ve standart hatalar benzer bir eğilim göstermiştir. Bu durum Carle'nin (2009) çalışmasındaki bulgularla benzerlik göstermektedir. Carle ağırlıklandırılmış model katsayıları ile ağırlıklandırılmamış model katsayıları arasında genel anlamda büyük farklılıklar olmadığını; buna karşın standart hataların karşılaştırılabilir olduğunu belirtmiştir. Ancak, HLM analizinde tüm makul değerlerin ağırlıklandırmalar ile birlikte kullanıldığı modellerin manidarlık bakımından daha tutucu olduğu, genel olarak açıklanan varyans oranını arttırdığı bundan dolayı hassas çalışmalarda kullanılması gerektiği söylenebilir.

Araştırma sonuçları hiyerarşik yapıdaki bir veri için HLM analizi kullanmanın çoklu doğrusal regresyon analizindekinden farklı sonuçlar verdiğini göstermektedir. Yuvalanmış bir veri için çoklu

doğrusal regresyon uygun ve yeterli olamayacağından HLM analizinin tercih edilmesi gerekmektedir. Bu sayede, farklı düzeydeki açıklayıcı olabilecek değişkenlerin tek başına ve birlikte etkileri gözlemlenebilecek, bağımlı değişkeni yordayan açıklayıcı değişkenler doğru ve güvenilir bir şekilde saptanabilecektir. Bu çalışmada sadece öğrenci düzeyi ağırlıklar kullanılmıştır. Benzer bir çalışma okul veya daha yüksek düzeye ait ağırlıklar kullanılarak tekrarlanabilir.

KAYNAKÇA

- Acar, T., & Öğretmen, T. (2012). Çok düzeyli istatistiksel yöntemler ile 2006 PISA fen bilimleri performansının incelenmesi. *Eğitim ve Bilim*, 37(163). Retrieved from <http://egitimvebilim.ted.org.tr/index.php/EB/article/download/1040/346>
- Adams, R. J., & Wu, M. L. (Eds.) (2002) *PISA 2000 technical report*. Paris: OECD Publications.
- Atar, B. (2010). Basit doğrusal regresyon analizi ile hiyerarşik doğrusal modeller analizinin karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 1(2), 78-84.
- Atar, H. Y., & Atar, B. (2012). Examining the effects of Turkish education reform on students' TIMSS 2007 science achievements. *Educational Sciences: Theory and Practice*, 12(4), 2632–2636.
- Beaton, A. E. (1987). *Implementing the new design*. (The NAEP 1983-84 technical report, Report No. 15-TR-20). Princeton, NJ: Educational Testing Service.
- Bock, R. D. & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of the EM algorithm. *Psychometrika* 46, 443-459.
- Bryk, A. S., & Raudenbush, S. W. (1988). Toward a more appropriate conceptualization of research on school effects: A three-level hierarchical linear model. *American Journal of Education* 97(1), 65-108.
- Bryk, A. S., & Raudenbush, S. W. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Bryk, A. S., Raudenbush, S. W., & Congdon, R. (2010). HLM7 for Windows [Computer software]. Chicago, IL: Scientific Software International, Inc.
- Carle, A. C. (2009). Fitting multilevel models in complex survey data with design weights: Recommendations. *BMC Medical Research Methodology*, 9(1), 1-13. doi: 10.1186/1471-2288-9-49
- Chowa, G. A., Masa, R. D., Ramos, Y., & Ansong, D. (2015). How do student and school characteristics influence youth academic achievement in Ghana? A hierarchical linear modelling of Ghana Youth Save baseline data. *International Journal of Educational Development*, 45, 129-140.
- Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). New York, NY: John Wiley and Sons.
- Fraenkel, J. R.; Wallen, N. E.; Hyun, H. H. (2012): *How to design and evaluate research in education* (8th Ed.). New York, NY: McGraw-Hill Humanities / Social Sciences/Languages.
- Gelman, A. (2006). Multilevel (hierarchical) modelling: What it can and cannot do. *Technometrics* 48(3), 432-435.
- Goldstein, H. (2011). *Multilevel statistical models* (Vol. 922). Oxford: John Wiley & Sons.
- International Business Machines Corp. (2015). IBM SPSS Statistics for Windows (Version 23.0) [Computer software]. Armonk, NY: IBM Corp.
- International Association for the Evaluation of Educational Achievement (2016), Help Manual for the IDB Analyzer. Hamburg, Germany: Retrieved from www.iea.nl/data
- Lohr, S. (2010). *Sampling: Design and analysis* (2nd edition). Boston, MA: Brooks / Cole.
- Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modelling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 1(3), 86-92. doi:10.1027/1614-2241.1.3.86
- Meinck, S. (2015). Computing sampling weights in large-scale assessments in education [Special issue]. *Survey Insights: Methods from the Field, Weighting: Practical Issues and 'How to' Approach*. Retrieved from <https://surveyinsights.org/?p=5353>
- Mislevy, R. J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika*, 56(2), 177–196.
- Mislevy, R. J. (1993). Should “multiple imputations” be treated as “multiple indicators”? *Psychometrika*, 58(1), 79–85.
- Organization for Economic Cooperation and Development (2009). Analyses with plausible values. In *PISA Data Analysis Manual: SPSS, Second Edition*, OECD Publishing. <http://dx.doi.org/10.1787/9789264056275-9-en> adresinden edinilmiştir.
- Organization for Economic Cooperation and Development (2014). *PISA 2012 technical report*. Paris: OECD.
- Organization for Economic Cooperation and Development (2017). *PISA 2015 Technical report*. Paris: OECD.

- Osborne, J. W. (2000). Advantages of hierarchical linear modeling. *Practical Assessment, Research & Evaluation*, 7(1), 1-3.
- Palardy, G. J. (2011). Review of HLM 7. *Social Science Computer Review*, 29(4), 515-520. doi: 10.1177/0894439311413437
- Rasch, G. (1960). *Studies in mathematical psychology: I. probabilistic models for some intelligence and attainment tests*. Oxford, England: Nielsen & Lydiche.
- Raudenbush, S. W. (1988). Educational applications of hierarchical linear models: A review. *Journal of Educational Statistics*, 13(2), 85-116.
- Raudenbush, S. W., & Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, 59(1), 1-17.
- Roberts, J. K. (2004). An introductory primer on multilevel and hierarchical linear modelling. *Learning Disabilities: A Contemporary Journal* 2, 30-38.
- Rubin, D. B. (1987). *Multiple imputations for non-response in surveys*. New York, NY: Wiley.
- Särndal, C., Swensson, B. & Wretman, J. (1992). *Model assisted survey sampling*. New York, NY: Springer-Verlag.
- Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis. An introduction to basic and advanced multilevel modeling*. London: Sage.
- Snijders, T., & Bosker, R. (2003). *Multilevel analysis: An introduction to basic and applied multilevel analysis*. Thousand Oaks, CA: Sage Publications.
- Stipek, D., & Valentino, R. A. (2015). Early childhood memory and attention as predictors of academic growth trajectories. *Journal of Educational Psychology*, 107(3), 771-788.
- Von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful. *IERI Monograph Series*, 2, 9-36.
- Warm, T. A. (1985). *Weighted maximum likelihood estimation of ability in item response theory with tests of finite length*. (Technical Report No. CGI-TR-85-08). Oklahoma, OK: Coast Guard Institute.
- Woltman, H., Feldstain, A., MacKay, J. C., & Rocchi, M. (2012). An introduction to hierarchical linear modeling. *Tutorials in Quantitative Methods for Psychology*, 8(1), 52-69.
- Wright, B.D., & Stone, M. H. (1979). *Best test design*. Chicago, IL: MESA Press.
- Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation*, 31(2), 114-128.
- Yang, H. (2013). The case for being automatic: Introducing the automatic linear modeling (LINEAR) procedure in SPSS statistics. *Multiple Linear Regression Viewpoints*, 39(2), 27-37.