



## PISA 2018 Okuma Becerileri Testinde Yer Alan Maddelerde Değişen Madde Fonksiyonunun İncelenmesi

Şerife Zeybekoğlu <sup>1</sup>, Ayşe Bilicioğlu Güneş <sup>2</sup>, Evrim Yalçın <sup>3</sup>

### Öz

Bu çalışmada, Uluslararası Öğrenci Değerlendirme Programı (PISA) 2018 uygulaması Türkiye örnekleme okuma becerileri alt testindeki maddeler üzerinde farklı yöntemler kullanılarak Değişen Madde Fonksiyonu (DMF) çalışmalarının yürütülmesi ve kullanılan yöntemlerin performanslarının karşılaştırılması amaçlanmaktadır. Yapılan analizlerde, bireyselleştirilmiş test deseni göz önünde bulundurularak temel bölüm, birinci aşama ve ikinci aşama bölümlerinde aynı demette yer alan madde paketleri ele alınmıştır. Temel Bölüm, birinci aşama ve ikinci aşama için sırasıyla ikinci paket (Core RC2), ikinci paket (Stage1 - R12H) ve üçüncü paket (Stage 2 - R23H) seçilmiştir. Bu pakette kısmi puanlanan 3 madde analiz dışı bırakılmış ve 1-0 şeklinde puanlanan 33 ortak madde analize dahil edilmiştir. Çalışmaya Türkiye’de bu madde paketini yanıtlayan 147 öğrenci alınmıştır. Öğrenci ve okul ölçekleri verilerinden alt problemlerde çalışılacak olan cinsiyet (ST004D01T), okulun bulunduğu yerleşim bölgesi (SC001Q01TA) değişkenleri ve ekonomik, sosyal ve kültürel düzey indeksi (ESCS) alınarak bilişsel testin maddeleri ile birleştirilmiştir. Verilerin analizine başlamadan önce veri seti düzenlenmiştir, kayıp veri ve uç değerler incelemesinin ardından kuramların varsayımları test edilmiştir. Çalışma kapsamında iki kategorili değişkenler için Mantel Haenszel (MH), Lojistik Regresyon (LR), SIBTEST ve Raju’nun Alan Ölçüleri yöntemleri; üç kategorili değişkenler için ise Genelleştirilmiş MH, Genelleştirilmiş LR ve Genelleştirilmiş Lord’un  $\chi^2$  yöntemleri kullanılmıştır. Cinsiyet değişkenine göre, MH, SIBTEST ve LR yönteminde sırasıyla iki, dört ve üç maddenin DMF içerdiğine ulaşılırken, Raju’nun Alan Ölçüleri yönteminde, işaretli alan testi sonucunda 17 ve işaretli alan testi sonucunda yedi maddenin DMF içerdiği tespit edilmiştir. ESCS değişkenine göre, MH ve LR yöntemlerinde sırasıyla iki ve bir maddenin DMF içerdiği tespit edilirken; Raju’nun Alan Ölçüleri yönteminde işaretli alan testi sonucunda 15 ve işaretli alan testi sonucunda sekiz maddenin DMF içerdiği belirlenmiştir. SIBTEST yönteminde ise DMF içeren madde tespit edilmemiştir.

### Anahtar Kelimeler

PISA 2018  
Okuma becerileri  
Değişen Madde Fonksiyonu (DMF)  
Cinsiyet  
Ekonomik, Sosyal ve Kültürel Düzey İndeksi (ESCS)  
Okulun bulunduğu yerleşim bölgesi

### Makale Hakkında

Gönderim Tarihi: 16.08.2022  
Kabul Tarihi: 25.05.2023  
Elektronik Yayın Tarihi: 03.11.2023

DOI: 10.15390/EB.2023.12123

<sup>1</sup> Ankara Üniversitesi, Eğitim Bilimleri Fakültesi, Eğitim Bilimleri Bölümü, Türkiye, [serifezeybekoglu79@gmail.com](mailto:serifezeybekoglu79@gmail.com)

<sup>2</sup> TED Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Türkiye, [ayse.bilicioğlu@tedu.edu.tr](mailto:ayse.bilicioглу@tedu.edu.tr)

<sup>3</sup> Ankara Üniversitesi, Eğitim Bilimleri Fakültesi, Eğitim Bilimleri Bölümü, Türkiye, [evrim0626@gmail.com](mailto:evrim0626@gmail.com)

Okulun bulunduğu yerleşim bölgesi değişkenine göre, Genelleştirilmiş MH, Genelleştirilmiş LR ve Genelleştirilmiş Lord'un  $\chi^2$  yönteminde sırasıyla bir, iki ve 28 maddenin DMF içerdiği belirlenmiştir. Araştırma sonucunda KTK ve MTK'ya dayalı yöntemler kendi içerisinde tutarlık gösterse de DMF düzeylerinde farklılık görülmüştür. MTK'ya dayalı yöntemlerde KTK'ya dayalı yöntemlere göre daha fazla madde DMF'li olarak tespit edilmiştir. Ek olarak Genelleştirilmiş MH ve Genelleştirilmiş LR yöntemlerinde ise benzer sonuçlar elde edilmiştir.

## Giriş

Okuma becerisi geçmişten günümüze önemli bir kavram olarak karşımıza çıkmaktadır. Eğitim ve öğretimin her aşamasında yer edinmiş olan okuma becerisi teknolojik gelişmeler ile birlikte evrilmeye başlamıştır. Başka bir ifadeyle, geçmiş yıllarda ihtiyaç duyulan okuma becerilerinin nitelik ve içeriği ile günümüzdeki farklı bir hal almıştır. Günümüzde okuma yalnızca yazılı kaynaklarla değil, elektronik kaynaklarla da yapılabilir hale gelmiştir. Bu durum farklı kaynakları kullanabilmeyi, belirsiz bir durum karşısında yönümüzü belirleyebilmeyi, algı ile gerçek durum arasındaki farklılıkları anlayabilmeyi içeren okuryazarlık kavramını karşımıza çıkartmıştır (Milli Eğitim Bakanlığı [MEB], 2019). Okuryazarlık kavramı, değerlendirmenin de bir parçası haline gelmiştir ve Türkiye'nin de düzenli olarak katılım gösterdiği geniş ölçekli bir test olan Uluslararası Öğrenci Değerlendirme Programı'nın (Programme for International Student Assessment-PISA) da en belirgin özelliği olmuştur. PISA gibi uluslararası platformlarda yapılan değerlendirme çalışmalarına katılarak, ülkeler, diğer ülkelere göre ne durumda olduklarını belirleyip kendi eğitim sistemlerini gözden geçirme imkânı bulmaktadır.

PISA matematik okuryazarlığı, fen okuryazarlığı ve okuma becerileri alanlarından oluşan ve her üç yılda bir döngüsel olarak bu alanlardan birini ağırlıklı alan olarak belirleyen bir araştırmadır. 2018 yılında gerçekleştirilen PISA uygulamasında ağırlıklı alan ise okuma becerileri olmuştur. Bu bağlamda PISA 2018'de okuma becerileri şu şekilde tanımlanmıştır (OECD, 2019): *“Okuma becerileri, kişinin hedeflerine ulaşmak, bilgi ve potansiyelini geliştirmek ve topluma katılmak amacıyla çeşitli şekillerde sunulan metinleri anlaması, kullanması, değerlendirmesi, ilişkilendirmesi ve metinler üzerine derinlemesine düşünmesidir.”*

Öğrencilerin okuma becerilerindeki başarısı diğer akademik alanlarda gösterdiği beceriler açısından da önemli görülmektedir. Bir öğrencinin okuma becerisindeki performansı düşük ise genel olarak diğer becerilerin kazanılmasında da zorluklar yaşayacağı düşünülmektedir (Geske ve Ozola, 2008). Bireylerin fen bilimleri ve matematik alanlarında başarılı olabilmeleri için öncelikle metni ve sembollerini iyi okuyup anlaması ve okuduklarını yorumlaması gerekmektedir. Bu yönüyle de okuduğunu anlama becerisi oldukça kritiktir. Rindermann ve Baumeister (2015) PISA uygulamalarındaki öğrenci başarılarının (fen ve matematik performansı dahil) yorumlanmasında okuma performansının dikkate alınmasının oldukça önemli olduğunu vurgularken; Akbaşı, Şahin ve Yaykiran (2016) PISA 2006, 2009 ve 2012 verilerinin yanı sıra öğrenciler ve öğretmenlerle yaptıkları ilgili çalışmalardan okuduğunu anlamının matematik ve fen başarısının önemli bir yordayıcısı olduğu sonucuna ulaşmıştır. Fuentes (1998) ise matematik ve okumanın birlikte gittiğini savunmuştur; diğer bir ifadeyle öğrencilerin matematik başarılarını artırmak için okuma becerilerinin de geliştirilmesi gerektiğini vurgulamıştır. Okuma becerisi ile matematik başarısı arasındaki ilişkiyi ortaya koyan daha pek çok sayıda çalışmaya rastlamak da mümkündür (Ding ve Homer, 2020; Erdem, 2016; Grimm, 2008; Lerkkanen, Rasku-Puttonen, Aunola ve Nurmi, 2005; Osterholm, 2005).

PISA uygulamalarında ölçme araçları İngilizce ve Fransızca olarak iki dilde hazırlanıp sınava katılacak ülkelere kendi ulusal dillerine çevirmeleri için gönderilmektedir. PISA kapsamında yer alan testler ve anketler farklı dillere çevrildiğinde formlar arasında eş değerliğin sağlanması gerekir. Eğer bir çalışma farklı dil ve kültürdeki bireyleri karşılaştırma amacı taşıyorsa, bu karşılaştırmanın anlamlı

olabilmesi için ölçme eş değerliğinin sağlanması gerekir. Uluslararası düzeyde yapılan uygulamalarda bireyler çeşitli özellikler bakımından farklılık göstermektedir. Sınav dilinin farklı olması, cinsiyet, sosyoekonomik düzey ve kültürel özellikler bireylerin sınav performansını etkilemektedir. Uluslararası düzeyde uygulanan ve ülkelerin eğitim sistemine yön veren bu uygulamanın minimum düzeyde hata içermesi gerekir. Bu sebeple, ülkeler arasında yapılan karşılaştırmaların anlamlı olabilmesi için ölçülen yapının gruptan bağımsız olması gerekir. Gruptan bağımsızlık, farklı alt gruplarda bulunan aynı yetenek düzeyindeki bireylerin puanlarının eşit olmasıdır (Osterlind, 1983). PISA uygulamalarında gruptan bağımsızlık koşulu sağlanamazsa karşılaştırmalar sonucunda ortaya çıkan farklılıkların gerçek farktan mı yoksa yapının gruptan gruba değişmesinden mi kaynaklandığı bilinemez. Bu sebeple, farklı dil ve kültürlerde uluslararası düzeyde yapılan PISA uygulamalarında sonuçlara hata karışması geçerlik sorunlarına ve madde yanlılığına sebep olabileceği için bu konuların değerlendirilmesi önem kazanmaktadır (Gök, Kabasakal ve Kelecioğlu, 2014).

Bir ölçme aracının ölçtüğü özelliklerin farklı gruplar için değişmez olmaması, test puanlarının belli bir gruba yönelik sistematik hata içermesi olarak tanımlanan yanlılık kavramına karşılık gelmektedir (Camilli ve Shepard, 1994). Ölçme aracının yanlılık içermesi, o ölçme aracından elde edilen verilerle alınan kararların, yapılan yorumların ve karşılaştırmaların geçerliğini sorgulanır hale getirmektedir. Yanlılık, ölçme aracından elde edilen sonuçların geçerliğine yönelik en büyük tehditlerden biridir (Clauser ve Mazor, 1998; Kristjansson, Aylesworth, Mcdowell ve Zumbo, 2005; Zumbo, 1999). Bir maddenin yanlı olması halinde, maddenin doğru yanıtlanması ölçülen yetenekten ziyade herhangi bir gruba ait olmasına bağlı olmaktadır (Osterlind, 1983). Bu bağlamda, bir test veya test maddesinin geçerlik şartını yerine getirebilmesi için maddenin yanlı olmaması önemli bir ölçüt olarak görülmektedir (Camilli ve Shepard, 1994).

Yanlılık çalışmaları, Değişen Madde Fonksiyonu (DMF) belirleme yöntemlerinin kullanılmasıyla başlayan ve uzman görüşünün alınması ile devam eden süreci içerir. Yanlılığın istatistiksel olarak manidarlığının görülebilmesi amacıyla DMF belirleme yöntemleri kullanılır. Ardından DMF içeren madde ya da maddelerdeki farklı fonksiyonlaşmanın kaynağının madde etkisi olarak adlandırılan gruplar arası gerçek farktan mı, yanlılıktan mı kaynaklandığının belirlenmesi amacıyla uzman görüşlerine başvurulur.

Literatürde DMF'nin belirlenmesi amacıyla kullanılacak çok sayıda yöntem bulunmaktadır. İki ya da çok kategorili değişkenlerin kullanılması ve iki ya da daha fazla sayıda grubun bulunması halinde kullanılacak DMF belirleme yöntemleri farklılaşmaktadır. Ancak, yöntemleri en genel haliyle Klasik Test Kuramını (KTK) ve Madde Tepki Kuramını (MTK) temel alan yöntemler olarak iki sınıfta toplamak mümkündür. KTK temelinde geliştirilen DMF belirleme yöntemlerinin en bilinenleri Varyans Analizi, Mantel-Haenszel (MH), SIBTEST, Dönüştürülmüş Madde Güçlüğü, Lojistik Regresyon (LR) olarak sıralanabilir. MTK temelinde geliştirilen yöntemlerden bazıları ise Lord'un  $\chi^2$ , Raju'nun Alan Ölçüleri ve Olabilirlik Oran Testi şeklinde sıralanabilir (Camilli ve Shepard, 1994; Hambleton, Swaminathan ve Rogers, 1991). Bazı MTK'ya dayalı DMF belirleme yöntemleri madde parametre kestirimlerine ya da madde tepki modelleri ile veriler arasındaki uyum iyiliğinin karşılaştırılmasına dayanmakta; bazıları iki çalışma grubundan elde edilen eğriler arasındaki farkı ölçmek veya anlamlılığı test etmek için istatistiksel testler geliştirmektedir (Thissen, Steinberg ve Wainer, 1993). DMF'nin tek biçimli (TB) ve tek biçimli olmayan (TBO) olmak üzere iki farklı türü bulunmaktadır. TB-DMF'de DMF etkisi yetenek düzeyi boyunca sabittir; TBO-DMF'de DMF etkisinin yetenek düzeyi boyunca büyüklüğünde ve yönünde farklılaşmalar görülür (Camilli ve Shepard, 1994). Başka bir deyişle, MTK terminolojisine göre, TB-DMF paralel madde karakteristik eğrisi kullanılarak gösterilirken, TBO-DMF paralel olmayan madde karakteristik eğrileri ile gösterilir. MTK tabanlı yöntemlerle TB-DMF'nin saptanmasında MTK Rasch modelleri veya bir parametrelili lojistik model ve uzantıları kullanılırken (Wright ve Masters, 1982); TBO-DMF'nin saptanmasında iki ya da üç parametrelili lojistik modeller ve bunların uzantıları kullanılır (Hambleton ve Swaminathan, 1985).

Literatürde geniş ölçekli testler üzerinde yürütülen DMF belirleme çalışmaları incelendiğinde, ağırlıklı olarak cinsiyet (Acar, 2011; Ateşok Devci, 2008; Bakan Kalaycıoğlu ve Kelecioğlu, 2011; Birjandi ve Amini, 2007; Çelik ve Özkan, 2020; Hamilton ve Snow, 1998; Öğretmen ve Doğan, 2004; Zenisky, Hambleton ve Robin, 2004), okul türü (Bakan Kalaycıoğlu, 2008; Gök, Kelecioğlu ve Doğan, 2010; Karakaya ve Kutlu, 2012; Kelecioğlu, Karabay ve Karabay, 2014; Şenferah, 2015; Yıldırım, 2015), bölge ve kültür (Ercikan ve Kim, 2005; Kabasakal ve Kelecioğlu, 2012; Ulutaş, 2012; Yurdugül ve Aşkar, 2004) temelli olarak yürütüldüğü görülmektedir. Öte yandan uluslararası düzeyde yapılan DMF çalışmalarında ise sıklıkla kültürler ve dilsel yanlılığın (Çet, 2006; Grisay ve Monseur, 2007; Gür, 2019; Le, 2006; Sırgancı, 2012; Uzun ve Gelbal, 2017) tespit edilmeye çalışıldığı görülmektedir. DMF çalışmalarında alınan örneklemin büyüklüğü, verilerin yapısı, maddelerin puanlanış biçimi, çalışmada kullanılan yöntemler, maddelerin DMF düzeylerinin araştırmalarda değişiklik göstermesine neden olabilmektedir. Bu sebeple, kullanılacak yönteme karar verebilmek için DMF belirleme yöntemlerinin karşılaştırılmasına dayanan çalışmalara ihtiyaç duyulmaktadır. Uluslararası alanda bu tür çalışmalara rastlanmakla birlikte, kullanılan yöntemlerin az sayıda olduğu ve çok sayıda yöntemin birbiriyle karşılaştırılmadığı görülmektedir. Bu çalışma, eğitim ve öğretimin temelinde yer alan okuma becerisinin PISA 2018 uygulamasından elde edilen verileri üzerinde yürütülmüş olması, farklı temellere dayanan birden çok DMF belirleme yönteminin bir arada kullanılması, bu yöntemleri kıyaslama fırsatı sunması ve uluslararası düzeyde kullanılan ölçme araçlarının geçerliğinin belirlenmesi yönleriyle önem taşımaktadır. Bunlara ek olarak, sıklıkla yanlılık kaynağı olarak karşımıza çıkan cinsiyet değişkeni ile görece sınırlı çalışmada söz konusu edilen yerleşim yeri (Yurdugül, 2003) ve sosyoekonomik düzey (Berberoğlu, 1995; Walzebug, 2014) değişkenlerinin de araştırmaya dahil edilmesi literatürdeki boşluğu dolduracak olması yönüyle önem arz etmektedir.

Bu çalışmada, PISA 2018 uygulaması Türkiye örneklemini okuma becerileri alt testindeki maddeler üzerinde farklı yöntemler kullanılarak DMF belirleme çalışmalarının yürütülmesi ve kullanılan yöntemlerin performanslarının karşılaştırılması amaçlanmaktadır. Bu amaçla, cinsiyet, ESCS ve okulun bulunduğu yerleşim bölgesi değişkenlerine göre DMF incelemesi yapılması planlanmıştır.

Çalışma kapsamında iki kategorili değişkenler için MH, LR, SIBTEST ve Raju'nun Alan Ölçüleri yöntemlerinin; üç kategorili değişkenler için ise Genelleştirilmiş MH, Genelleştirilmiş LR ve Genelleştirilmiş Lord'un  $\chi^2$  yöntemlerinin kullanılması planlanmıştır.

Çalışmanın amacı doğrultusunda oluşturulan alt problemlere aşağıda yer verilmiştir.

1. PISA 2018 uygulamasında okuma becerileri alt testinde yer alan maddeler MH, LR, SIBTEST ve Raju'nun Alan Ölçüleri yöntemleriyle incelendiğinde cinsiyet değişkenine göre DMF içermekte midir?
2. PISA 2018 uygulamasında okuma becerileri alt testinde yer alan maddeler MH, LR, SIBTEST ve Raju'nun Alan Ölçüleri yöntemleriyle incelendiğinde ESCS değişkenine göre DMF içermekte midir?
3. PISA 2018 uygulamasında okuma becerileri alt testinde yer alan maddeler Genelleştirilmiş MH, Genelleştirilmiş LR ve Genelleştirilmiş Lord'un  $\chi^2$  yöntemleriyle incelendiğinde okulun bulunduğu yerleşim bölgesi değişkenine göre DMF içermekte midir?

## Yöntem

### *Araştırma Modeli*

Araştırma kapsamında, PISA 2018 Türkiye örneklemini okuma becerileri alt testindeki maddelerin cinsiyet, ESCS ve okulun bulunduğu yerleşim bölgesi değişkenlerine göre DMF içerip içermediği incelenmiştir. Bu amaç doğrultusunda, çalışma betimsel bir araştırma özelliği taşımaktadır. Araştırmada PISA 2018 okuma becerileri alt testindeki maddeler farklı yöntemlerle DMF analizine tabi tutulmuş ve sonuçlar betimlenmiştir. Betimsel araştırmalar var olan durumun eksiksiz ve özenli bir şekilde tanımlanmasını sağlar (Büyüköztürk, Çakmak, Akgün, Karadeniz ve Demirel, 2009).

### **Çalışma Grubu**

PISA uygulamalarında tabakalı seçkisiz örnekleme yöntemiyle okul örnekleme belirlenmektedir. PISA 2018 uygulamasında Türkiye örnekleme belirlenirken okullar için kullanılan tabakalar İstatistikî Bölge Birimleri Sınıflaması (İBBS) Düzey 1, okul türü, okulun idari biçimi, okulun konumu ve cinsiyettir. Okullar belirlendikten sonra bu okullarda uygulamaya katılacak öğrenciler seçkisiz bir şekilde belirlenmiştir. Türkiye’de PISA 2018 uygulamasına İBBS Düzey 1’e göre 12 bölgeyi temsilen 186 okul ve 6890 öğrenci katılmıştır (MEB, 2019).

Bu çalışmada yapılan analizlerde, bireyselleştirilmiş test deseni göz önünde bulundurularak temel bölüm, birinci aşama ve ikinci aşama bölümlerinde aynı demette yer alan madde paketleri ele alınmıştır. Temel Bölüm, birinci aşama ve ikinci aşama için sırasıyla ikinci paket (Core RC2), ikinci paket (Stage1 - R12H) ve üçüncü paket (Stage 2 - R23H) seçilmiştir. Çalışmaya Türkiye’de bu madde paketini yanıtlayan 147 öğrenci dahil edilmiştir. Çalışmaya dahil edilen öğrenciler bu maddeleri aynı sıra ile görmüşlerdir.

Yapılan çalışmalarda örneklem büyüklüğü arttıkça DMF belirleme yöntemlerinin gücünün de arttığı belirlenmiştir (Narayanan ve Swaminathan, 1994; Rogers ve Swaminathan, 1993). Zieky (1993) DMF istatistiklerini hesaplamak için her grupta en az 200, toplamda ise 400 kişinin olması gerektiğini belirtmiştir. Bu çalışma ise yalnızca Türkiye örnekleme üzerinde yürütülmüştür. Örnekleme farklı madde paketlerini yanıtlayan öğrenci sayıları incelendiğinde, belirtilen örneklem sayısına ulaşamamış ve bu nedenle 147 öğrenciyle çalışma yürütülmüştür. Belzak (2020) birçok araştırmacının küçük örneklemde DMF tespitinde başarısız olduğunu çünkü yaygın olarak kullanılan yöntemlerin DMF’yi tespit edebildiğine dair çok az kanıt olduğunu belirtmiştir. Yapmış olduğu çalışmada, daha az karmaşık modeller kullanılarak 50-100 (her grupta 25-50) kadar düşük örneklem boyutlarında orta düzeyde DMF tespit edilebildiğini göstermiştir. Bu örneklem boyutu araştırmanın bir sınırlılığı olarak görülebilir.

### **Veri Toplama Araçları**

PISA 2018 uygulamasına 79 ülke katılmış, katılan ülkelerin 70’inde hem bilişsel test hem de anket uygulamaları bilgisayar tabanlı olarak; kalan 9 ülkede ise kâğıt-kalem testi olarak yapılmıştır (OECD, 2019). Türkiye’de ise PISA 2018 uygulaması bilgisayar tabanlı olarak gerçekleştirilmiştir.

PISA 2018 uygulamasına ait veri seti OECD PISA internet sitesinden indirilmiş, bu veri setinden Türkiye örnekleme seçilmiştir. PISA 2018 döngüsündeki bilişsel testlerde okuma becerileri maddelerinin sayısı çok aşamalı uyarlanabilir tasarım nedeniyle önceki döngülere göre çok daha fazladır. Okuma becerileri testi 245 maddeye ek olarak 65 akıcı cümle içermektedir.

PISA 2018 uygulamasında öğrenci başarısını daha doğru ölçebilmek amacıyla bireyselleştirilmiş test kullanılmıştır. PISA’nın önceki uygulamalarında kitapçıklardaki soruların her birinin yeri önceden belirlenmiştir, yani bu kitapçıklardaki sorular sabittir. Fakat PISA 2018 okuma becerileri testinde dinamik bir yapı geliştirilmiş ve sorular öğrencinin önceki sorulara verdiği cevaplara göre belirlenmiştir (OECD, 2019). Bu doğrultuda yapılandırılan PISA 2018 okuma becerilerindeki sorular Temel, birinci aşama ve ikinci aşama şeklinde üç aşamadan oluşmaktadır.

İlk olarak öğrenciler 7-10 soru içeren Temel Bölümdeki soruları cevaplamışlardır. Bu bölümdeki sorular genel olarak otomatik olarak puanlanabilmektedir. Öğrencinin bu bölümdeki doğru cevap sayısına göre başarısı düşük, orta ve yüksek olarak sınıflandırılmıştır (OECD, 2019). Temel bölümdeki sorular güçlük bakımından aralarında anlamlı fark olmayacak şekilde hazırlanmıştır. Birinci aşama ve ikinci aşamadaki sorular ise kısmen kolay ve kısmen zor olarak iki farklı düzeyde hazırlanmıştır. Öğrencinin temel bölümdeki başarısına göre 1. aşama soruları belirlenmiştir. Daha sonra, hem temel hem de birinci aşamadaki başarısına göre, ikinci aşama soruları belirlenmiştir. Standart bilgisayar tabanlı değerlendirmede 64 farklı madde paketi belirlenmiş ve öğrencilerin %75’ine uygulanmıştır. Alternatif bilgisayar tabanlı değerlendirmede 128 farklı madde paketi belirlenmiş ve öğrencilerin %25’ine uygulanmıştır.

Bu çalışmada standart bilgisayar tabanlı değerlendirmedeki 64 farklı madde paketinden biri seçilmiştir. Bu pakette kısmi puanlanan 3 madde analiz dışı bırakılmış ve 1-0 şeklinde puanlanan 33 ortak madde analize dahil edilmiştir. Bu maddeler Tablo 1’de gösterilmiştir.

**Tablo 1.** Ortak Maddeler ve Üniteleri

Ortak Maddeler	Üniteler
CR545Q02S	Machu Picchu
DR545Q04C	Machu Picchu
CR545Q06S	Machu Picchu
CR545Q07S	Machu Picchu
CR424Q02S	Fair Trade
CR424Q03S	Fair Trade
CR424Q07S	Fair Trade
CR404Q03S	Sleep
CR404Q06S	Sleep
CR404Q07S	Sleep
DR404Q10AC	Sleep
DR404Q10BC	Sleep
CR558Q02S	Microwave Ovens
DR558Q12C	Microwave Ovens
DR558Q04C	Microwave Ovens
CR558Q06S	Microwave Ovens
CR558Q09S	Microwave Ovens
CR437Q01S	Narcissus
DR437Q07C	Narcissus
CR437Q06S	Narcissus
CR543Q01S	Alfred Nobel
CR543Q03S	Alfred Nobel
CR543Q04S	Alfred Nobel
CR543Q09S	Alfred Nobel
CR543Q10S	Alfred Nobel
CR543Q13S	Alfred Nobel
DR543Q15C	Alfred Nobel
DR566Q03C	The Skellig Rocks
CR566Q04S	The Skellig Rocks
CR566Q05S	The Skellig Rocks
CR566Q14S	The Skellig Rocks
CR566Q06S	The Skellig Rocks
DR566Q12C	The Skellig Rocks

Öğrenci ve okul ölçekleri verilerinden alt problemlerde çalışılacak olan cinsiyet (ST004D01T), ESCS ve okulun bulunduğu yerleşim bölgesi (SC001Q01TA) değişkenleri alınarak bilişsel testin maddeleri ile birleştirilmiştir.

ESCS değişkeninin ortalaması alınmış, ortalamanın altında olanlar düşük ESCS, ortalamanın üstünde olanlar yüksek ESCS olacak şekilde tanımlanarak değişken kategorik hale getirilmiştir.

Okulun bulunduğu yerleşim bölgesi değişkeni beş kategorili olmasına rağmen bazı alt grupların örneklem büyüklüğü çok düşüktür. Bu sebeple, kategoriler birleştirilmiş ve 3 kategoriye düşürülmüştür. Nüfusu 100 binden az olanlar ilçe, nüfusu 100 bin ile 1 milyon arasında olanlar şehir, nüfusu 1 milyondan fazla olanlar ise büyükşehir olacak şekilde kategorilendirilmiştir.

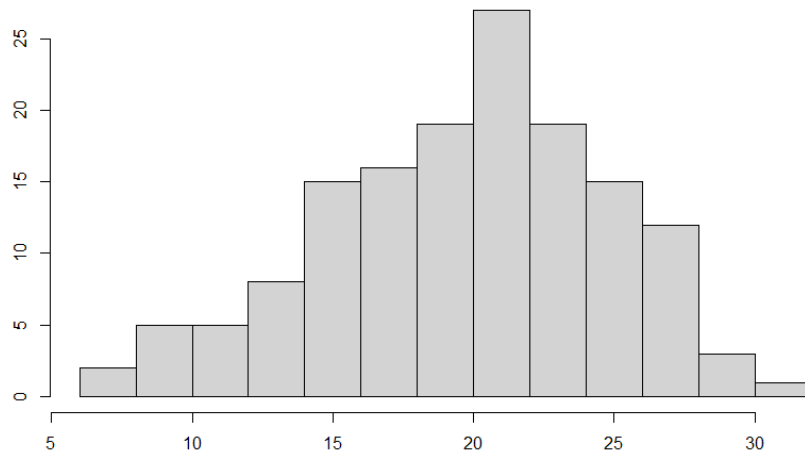
### Verilerin Analizi

Verilerin analizine başlamadan önce veri seti düzenlenmiştir. Veri setinde kayıp veri ve uç değerler incelenmiştir. Yapılan kayıp veri analizi sonucunda, Little ve Rubin (2002) tarafından yapılan sınıflandırmaya göre veri seti içerisinde yer alan kayıp verilerin rastlantısal olarak dağıldığı görülmüştür. Bu noktada, kayıp verileri gidermek amacıyla, kayıp verilerle baş etme yöntemlerinden çoklu atama (multiple imputation) yöntemi kullanılarak kayıp veriler giderilmiş ve her bir örneklem için tam bir veri seti elde edilmiştir. Daha sonra veri setindeki uç değerler incelenmiştir. Tek değişkenli uç değerlerin belirlenmesi için toplam puanlar z standart puanına dönüştürülerek analiz edilmiştir. Z puanı gözlenen değişkenin ortalama değere kaç standart sapma kadar uzakta olduğunu gösterir. Z puanı değerinin  $\pm 3$  aralığını aşması durumunda veri uç değer olarak değerlendirilir (Tabachnick ve Fidell, 2015). Bu çalışmada yapılan tek değişkenli uç değer analizi sonucunda uç değer tespit edilmemiştir. Kayıp veri ve uç değer analizlerinden sonra dağılımın normalliği incelenmiştir. Elde edilen sonuçlar Tablo 2 ve Şekil 1’de verilmiştir.

**Tablo 2.** Merkezi Eğilim Ölçüleri

Mod	Medyan	Aritmetik Ortalama	Çarpıklık Katsayısı	Basıklık Katsayısı
22.00	21.00	20.21	-0.351	-0.244

Merkezi eğilim ölçülerinden mod ( $X_{\text{mod}} = 22.00$ ), medyan ( $X_{\text{ort}} = 21.00$ ) ve aritmetik ortalama ( $\bar{X} = 20.21$ ) değerleri incelenmiş, çarpıklık katsayısı ( $K_y = -0.351$ ) ve basıklık katsayısının ( $B_s = -0.244$ )  $\pm 1$  sınırları arasında olduğu belirlenmiştir.



**Şekil 1.** Histogram Grafiği

Dağılımın histogram grafiği de incelendiğinde normale yakın bir dağılım olduğu tespit edilmiştir. 147 kişilik veriyle çalışmaya başlama kararı alınmıştır. Bu verinin cinsiyet, ESCS ve okulun bulunduğu yerleşim bölgesi kategorik değişkenlerine dağılımı Tablo 3’te görülmektedir.

**Tablo 3.** Kategorik Değişkenler

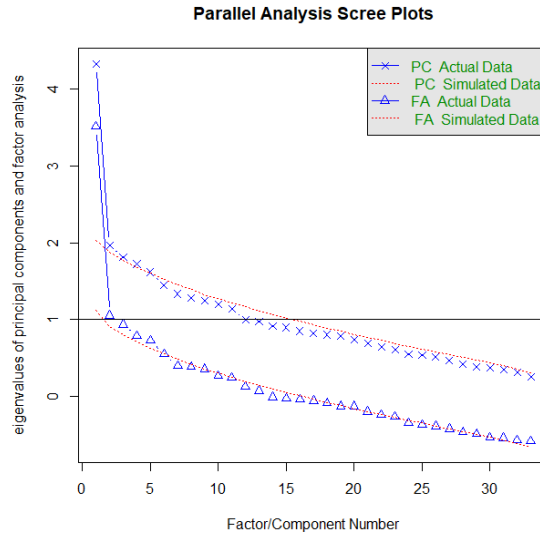
Cinsiyet		ESCS		Okulun Bulunduğu Yer		
Kadın	Erkek	Düşük	Yüksek	Kasaba	Şehir	Büyük Şehir
69	77	77	70	45	40	62

Bu çalışmada hem KTK’ya (MH, SIBTEST, LR, Genelleştirilmiş MH ve Genelleştirilmiş LR) hem de MTK’ya (Raju’nun Alan Ölçüleri ve Genelleştirilmiş Lord’un  $\chi^2$ ) dayalı yöntemlerle DMF incelemesi yapılmıştır. Öncelikle varsayımlar incelenmiş daha sonra DMF analizi gerçekleştirilmiştir. Çünkü KTK’ya dayalı DMF yöntemleri “tek boyutluluk” varsayımı gerektirirken, MTK’ya dayalı DMF yöntemleri ise MTK modelleri kullanıldığı için “tek boyutluluk, yerel bağımsızlık ve model veri uyumu” varsayımları gerektirmektedir.

### Varsayımların İncelenmesi

#### Tek boyutluluk

Tek boyutluluk incelemesi yapılmış ve paralel analiz sonucunda oluşan “Yamaç Eğim Grafiği” Şekil 2’de verilmiştir.



Şekil 2. Yamaç Eğim Grafiği

Yapılan inceleme sonucunda bu varsayımın karşılandığı ve yapının tek boyutlu olduğu belirlenmiştir.

#### Yerel Bağımsızlık

Verinin yerel bağımsızlığı Yen’in Q3 istatistiği ile incelenmiştir. Elde edilen sonuçlar Tablo 4’te görülmektedir.

Tablo 4. Yen’in Q3 istatistiği

M	SD	Min	%10	%25	%50	%75	%90	Max
-0.030	0.093	-0.312	-0.156	-0.094	-0.024	0.035	0.080	0.293

Yen’in Q3 istatistiği sonucunda artık korelasyonun -0.030 olduğu tespit edilmiştir. Bu korelasyon 0.20’nin altında olduğu için yerel bağımsızlığın sağlandığı belirlenmiştir (DeMars, 2016).

#### Model Veri Uyumu

Araştırmadaki veri 1-0 şeklinde ikili puanlandığı için Bir Parametrelili Lojistik Model (1 PL), İki Parametrelili Lojistik Model (2 PL) ve Üç Parametrelili Lojistik Model (3 PL) kestirimleri yapılmıştır. Elde edilen sonuçlar Tablo 5 ve Tablo 6’da görülmektedir.

Tablo 5. 1PL ve 2PL Model Arasında Yapılan ANOVA Sonuçları

	AIC	BIC	log.Lik	LRT	df	p
1 PL	5656.70	5758.38	-2794.35			
2 PL	5627.26	5824.63	-2747.63	93.44	32	<0.001

Tablo 6. 2PL ve 3PL Model Arasında Yapılan ANOVA Sonuçları

	AIC	BIC	log.Lik	LRT	df	p
2 PL	5627.26	5824.63	-2747.63			
3 PL	5670.01	5966.06	-2736.00	23.25	33	0.896



Tablo 5'te 1 PL ve 2 PL model arasında yapılan ANOVA sonucunda modeller arasında anlamlı bir fark olduğu ve 2PL modelin AIC ve log.Lik uyum indekslerinin daha düşük olduğu görülmüştür ve bu durum daha iyi uyuma işarettir. Tablo 6'da 2 PL ve 3 PL model arasında yapılan karşılaştırmada modeller arasında anlamlı bir fark görülmemiştir. Bu durumda iki kategorili modeller arasında yapılan karşılaştırmada 2PL modelin daha iyi uyum verdiği belirlenmiştir.

Çalışma sürecinde DMF incelemeleri R 4.0.3 programında "difR", "lrm", "psych", "sirt" ve "ShinyItemAnalysis" paketleri kullanılarak gerçekleştirilmiştir. Analizlerde hesaplanan DMF'lerin etki büyüklükleri için kesme noktaları aşağıda verilmiştir.

#### **Mantel-Haenszel (MH)**

MH yöntemine dayalı DMF belirlemede Zieky (1993) tarafından  $\Delta_{MH}$  değerini dikkate alarak bir sınıflama sistemi düzenlenmiştir.

**Tablo 7.** Zieky (1993)'nin Düzenlediği Sınıflama Sistemi

Değer Aralığı	DMF Düzeyi	Açıklama
$ \Delta_{MH}  < 1.0$	A	DMF yoktur veya ihmal edilebilir düzeydedir
$1.0 \leq  \Delta_{MH}  < 1.5$	B	DMF orta düzeydedir
$ \Delta_{MH}  \geq 1.5$	C	DMF yüksek düzeydedir

$\Delta_{MH}$  değerinin pozitif olması maddenin odak grup lehine, negatif olması ise referans grup lehine DMF içerdiği ve sifıra eşit olması ise DMF içermediği şeklinde yorumlanmaktadır (De Ayala, 2009).

#### **Genelleştirilmiş MH**

Bu yöntem MH yönteminin bir uzantısı olarak çok kategorili veriler için alternatif olarak geliştirilmiştir. MH yönteminde kategoriler sıralı olarak ele alınırken, Genelleştirilmiş MH yönteminde kategoriler sınıflamalı olarak ele alınır (Wang ve Su, 2004). Genelleştirilmiş MH yönteminin birçok avantajı bulunmasına rağmen DMF türü, etki büyüklüğü ve yönü hakkında detaylı bilgi vermemektedir (Fidalgo ve Scalón, 2012).

#### **SIBTEST**

SIBTEST yöntemine dayalı DMF belirlemede Roussos ve Stout (1996) tarafından  $\beta$  indeksini yorumlayabilmek için bir sınıflama sistemi düzenlenmiştir.

**Tablo 8.** Roussos ve Stout (1996)'un Düzenlediği  $\beta$  ölçütleri

Değer Aralığı	DMF Düzeyi	Açıklama
$ \beta  < 0.059$	A	DMF yoktur veya ihmal edilebilir düzeydedir
$0.059 \leq  \beta  < 0.088$	B	DMF orta düzeydedir
$ \beta  \geq 0.088$	C	DMF yüksek düzeydedir

SIBTEST analizi sonucunda oluşan  $\beta$  indeksinin pozitif olması referans grup lehine, negatif olması ise odak grup lehine DMF içerdiği şeklinde yorumlanmaktadır.

#### **Lojistik Regresyon (LR)**

LR yönteminde değişkenler sırasıyla modele eklenir. Model-1'de toplam puan modele alınırken; Model 2'de grup değişkeni de modele eklenir ve Model 3'te de etkileşim değişkeni grup ve toplam puan değişkeninin yanına eklenir. Oluşturulan modellerle maddenin TB ya da TBO-DMF içerip içermediği tespit edilebilir. Her bir modeldeki  $R^2$  değerleri arasındaki fark TB ve TBO-DMF değerlerini verir. TB ve TBO-DMF belirlenirken yüksek olan  $R^2$  değeri dikkate alınır. Model-3 ve Model-1'den elde edilen Nagelkerke  $R^2$  değerleri karşılaştırılarak  $\Delta R^2$  değeri elde edilir ve etki büyüklüğü hesaplanır. LR için Zumbo ve Thomas (1996) ve Jodoin ve Gierl (2001) tarafından bir sınıflama sistemi düzenlenmiştir.

**Tablo 9.** Zumbo ve Thomas (1996) ile Jodoin ve Gierl (2001)'in Düzenlediği Sınıflama Sistemi

Değer Aralığı		DMF Düzeyi	Açıklama
Zumbo ve Thomas	Jodoin ve Gierl		
$\Delta R^2 < 0.13$	$\Delta R^2 < 0.035$	A	DMF yoktur veya ihmal edilebilir düzeydedir
$0.13 \leq \Delta R^2 < 0.26$	$0.035 \leq \Delta R^2 < 0.070$	B	DMF orta düzeydedir
$\Delta R^2 \geq 0.26$	$\Delta R^2 \geq 0.070$	C	DMF yüksek düzeydedir

**Genelleştirilmiş LR**

Genelleştirilmiş Lojistik Regresyon (GLR), LR yönteminin genişletilmiş halidir. Bu yöntem ile çoklu gruplar arasında hem tek biçimli hem de tek biçimli olmayan DMF kestirimleri yapılabilmektedir (Magis, Raîche, Béland ve Gérard, 2011). Eşleşme kriteri bulunan bir model içeren bu yöntemde grup üyeliği ve grup-puan etkileşimi ile ilgili parametrelerin istatistiksel olarak anlamlılığı olasılık-oran testi aracılığıyla değerlendirilir. Madde yanıtları ile grup üyeliği arasında ilişki olması durumunda, maddenin DMF içerdiğine karar verilir (Magis vd., 2011).

**Raju'nun Alan Ölçüleri**

DMF yöntemlerinden Raju'nun Alan Ölçülerine göre analiz yapılırken işaretli alan indeksi ve işaretli alan indeksi hesaplanır. Bu indeksler negatif ise odak grup lehine, pozitif ise referans grup lehine DMF olduğunun göstergesidir. TBO-DMF'den söz edebilmek için işaretli alan indeks değeri işaretli alan indeks değerinden küçük olmalı, TB-DMF'den söz edebilmek için ise işaretli alan indeks değeri işaretli alan indeks değerinden büyük veya eşit olmalıdır. Raju'nun Diferansiyel Test ve Madde Fonksiyonunun (DFIT) en büyük dezavantajı etki büyüklüğünden yoksun olmasıdır. Oshima ve Wright (2015), DFIT'e MH tabanlı olarak etki büyüklüğünü tanımlamak için bir yaklaşım önermiştir. Bu yaklaşımda hesaplanmış  $\Delta_{MH}$  değeri dikkate alınır ve değer yaklaşık  $K=-15$  sabitine bölünerek  $\beta_u$  değeri bulunur (Shealy ve Stout, 1993). 1 PL ve 2 PL modellerde K sabiti yaklaşık -15 değerini alırken 3 PL modellerde yaklaşık -17 değerini alır.

$$\beta_u = \Delta_{MH}/K$$

Elde edilen  $\beta_u$  değerinin karesinin alınması ile NCDIF (Non-compensatory Differential Item Functioning) değeri hesaplanır.

$$NCDIF = (\Delta_{MH}/K)^2$$

Oshima ve Wright (2015)'a göre hesaplanan NCDIF değerleri için önerilen sınıflama aşağıda verilmiştir.

**Tablo 10.** Oshima ve Wright (2015)'in Düzenlediği Sınıflama Sistemi

Değer Aralığı	DMF Düzeyi	Açıklama
$NCDIF < 0.003$	A	DMF yoktur veya ihmal edilebilir düzeydedir
$0.003 \leq NCDIF < 0.008$	B	DMF orta düzeydedir
$NCDIF \geq 0.008$	C	DMF yüksek düzeydedir

**Genelleştirilmiş Lord'un  $\chi^2$** 

Lord'un  $\chi^2$  istatistiği test edilirken alt gruplar için madde parametreleri ile kovaryans hesaplanıp kestirilen parametreler ortak bir ölçeğe dönüştürülür. Böylece ölçeklenmiş parametre ve kovaryans değerleri kullanılarak Lord'un  $\chi^2$  istatistiği hesaplanabilmektedir. Sonrasında elde edilen değerler serbestlik derecesine göre  $\chi^2$  tablosundaki kritik değer ile kıyaslanarak DMF'nin varlığı incelenmektedir (Camilli ve Shepard, 1994). Kim, Cohen ve Park (1995) tarafından Lord'un  $\chi^2$  tekniği ikiden fazla grupta kullanılmak üzere genelleştirilerek Genelleştirilmiş Lord'un  $\chi^2$  tekniğini oluşturmuşlardır.

## Bulgular

### Birinci Alt Probleme İlişkin Bulgular

PISA 2018 okuma becerileri alt testinde yer alan maddelerin cinsiyet değişkenine göre DMF içerip içermediği MH, SIBTEST, LR ve Raju'nun Alan Ölçüleri yöntemleriyle test edilmiştir.

MH yöntemine göre DMF içeren maddeler Tablo 11'de görülmektedir.

**Tablo 11.** Cinsiyet Değişkeni İçin MH Sonuçları

Maddeler	$\alpha_{MH}$	$\chi^2$	p	$\Delta_{MH}$	DMF Düzeyi	Hangi Grup Lehine
DR558Q12C	2.6464	2.4669	0.0136*	-2.2870	C	Referans grup
CR566Q05S	2.5394	2.3016	0.0214*	-2.1900	C	Referans grup

Not. Referans grup: Erkek (N=78), Odak grup: Kadın (N=69); \*p<.05

Tablo 11'de p değerleri manidar olan maddelerin  $\Delta_{MH}$  değerleri incelenmiş olup Zieky (1993)'nin MH yöntemi için sınıfladığı  $\Delta_{MH}$  eşik değerleriyle karşılaştırılarak maddelerin hangi düzeyde DMF içerdiği belirlenmiştir. Buna göre iki maddenin yüksek düzeyde (C) DMF içerdiği tespit edilmiştir.

$\Delta_{MH}$  değerinin pozitif ya da negatif olmasına göre DMF içeren maddenin hangi grubun lehine olduğu belirlenmiştir. DR558Q12C ve CR566Q05S maddelerinin referans grup lehine (erkekler) DMF içerdiği tespit edilmiştir.

SIBTEST yöntemine göre DMF içeren maddeler Tablo 12'de görülmektedir.

**Tablo 12.** Cinsiyet Değişkeni İçin SIBTEST Sonuçları

Maddeler	$\beta$	SH	$\chi^2$	p	DMF Düzeyi	Hangi Grup Lehine
CR545Q06S	0.5413	0.2365	5.2376	0.0221*	C	Referans grup
CR558Q09S	-0.5417	0.2559	4.4817	0.0343*	C	Odak grup
DR437Q07C	-0.5101	0.2324	4.8169	0.0282*	C	Odak grup
CR566Q14S	0.5958	0.2108	7.9870	0.0047**	C	Referans grup

Not. Referans grup: Erkek (N=78), Odak grup: Kadın (N=69); \*p<.05 \*\*p<.01

Tablo 12'de p değerleri manidar olan maddelerin  $\beta$  değerleri Roussos ve Stout (1996)'un önerdiği  $\beta$  indeksi etki büyüklüğü yorumlama ölçütleriyle karşılaştırılmış ve maddelerin hangi düzeyde DMF içerdiği belirlenmiştir. Buna göre dört maddenin yüksek düzeyde (C) DMF içerdiği tespit edilmiştir.

$\beta$  değerinin pozitif ya da negatif olmasına göre DMF içeren maddenin hangi grubun lehine olduğu belirlenmiştir. CR545Q06S ve CR566Q14S maddelerinin referans grup (erkekler) lehine; CR558Q09S ve DR437Q07C maddelerinin odak grup (kadınlar) lehine DMF içerdiği tespit edilmiştir.

LR yöntemine göre DMF içeren maddeler Tablo 13'te görülmektedir.

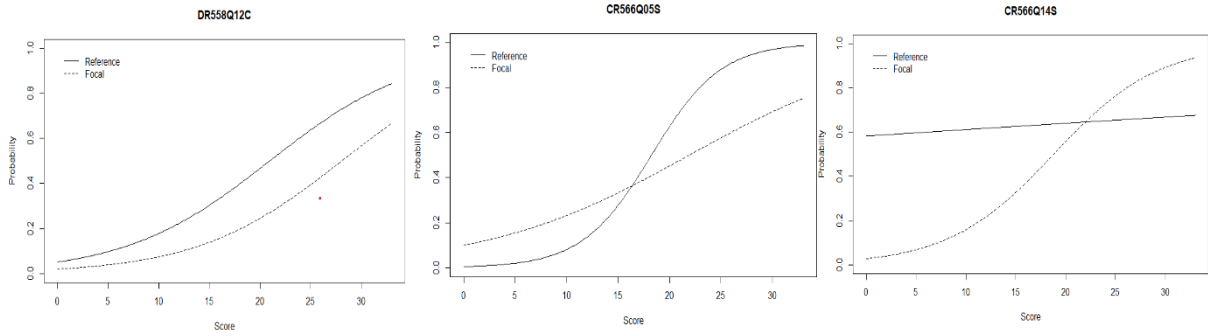
**Tablo 13.** Cinsiyet Değişkeni İçin LR Sonuçları

Maddeler	TB DMF $R^2$	TBO DMF $R^2$	$\chi^2$	p	$R^2$	DMF Düzeyi Jodoin ve Gierl	DMF Biçimi
DR558Q12C	0.0612	0.0089	7.0653	0.0079**	0.0612	B	TB
CR566Q05S	0.0297	0.0388	4.8815	0.0271*	0.0388	B	TBO
CR566Q14S	0.0092	0.0497	5.3366	0.0209*	0.0497	B	TBO

Not. Referans grup: Erkek (N=78), Odak grup: Kadın (N=69); \*p<.05 \*\*p<.01

Tablo 13'te p değeri manidar olan maddelerin Nagelkerke  $R^2$  değerleri incelenmiş, Jodoin ve Gierl (2001)'in LR yöntemi için düzenlediği sınıflama sistemi kullanılarak maddelerin hangi düzeyde DMF içerdiği belirlenmiştir. Buna göre üç maddenin orta düzeyde (B) DMF içerdiği tespit edilmiştir.

DMF'nin biçimini belirlemek için TB- $R^2$  ve TBO- $R^2$  değerlerinin büyüklükleri karşılaştırılmıştır. Ardından, madde düzeyinde grafiksel incelemeler yapılmıştır. Bu maddelere ait Madde Karakteristik Eğrileri (MKE) Şekil 3'te verilmiştir.



Şekil 3. Madde Karakteristik Eğrileri (Cinsiyet)

DMF tespit edilen maddenin grafiksel incelemesi sonucunda CR566Q05S ve CR566Q14S maddelerinin TBO-DMF ve DR558Q12C maddesinin TB-DMF içerdiği tespit edilmiştir. DR558Q12C maddesinde referans grup (erkekler) lehine DMF belirlenmiştir.

Raju'nun Alan Ölçüleri yöntemine göre DMF içeren maddeler Tablo 14'te görülmektedir.

Tablo 14. Cinsiyet Değişkeni İçin Raju'nun Alan Ölçüleri Sonuçları

Maddeler	İşaretsiz Alan		İşaretili Alan		$\Delta_{MH}$	NCDIF = ( $\Delta_{MH} / K$ ) <sup>2</sup>	DMF Düzeyi	DMF Biçimi
	Z	p	Z	p				
DR545Q04C	2.0468	0.0407*	-1.9710	0.0487*	-0.2212	0.000217	A	TBO
CR545Q06S	2.6087	0.0091**	2.4553	0.0141*	-0.8812	0.003451	B	TBO
CR545Q07S	2.6412	0.0083**	2.2503	0.0244*	-0.6329	0.00178	A	TBO
CR404Q06S	2.0720	0.0383*	-0.7329	0.4636	0.1535	0.000105	A	TBO
CR404Q07S	2.2147	0.0268*	-2.3455	0.0190*	0.7259	0.002342	A	TB
DR404Q10AC	2.9792	0.0029**	-1.2420	0.2142	0.3424	0.000521	A	TBO
DR404Q10BC	3.4343	0.0006***	-3.3195	0.0009***	0.5602	0.001395	A	TBO
DR558Q04C	2.5646	0.0103*	2.5053	0.0122*	0.6033	0.001618	A	TBO
CR558Q06S	2.2707	0.0232*	1.6120	0.1070	0.7825	0.002721	A	TBO
CR558Q09S	2.1716	0.0299*	1.3388	0.1806	0.4701	0.000982	A	TBO
CR543Q03S	2.8965	0.0038**	2.8860	0.0039**	0.8620	0.003302	B	TBO
CR543Q04S	2.0638	0.0390*	1.5308	0.1258	0.9459	0.003977	B	TBO
CR543Q09S	2.4450	0.0145*	-1.7340	0.0829	0.6012	0.001606	A	TBO
CR543Q13S	2.8596	0.0042**	0.0080	0.9936	0.4772	0.001012	A	TBO
CR566Q05S	2.5916	0.0096**	0.3390	0.7346	-2.1900	0.021316	A	TBO
CR566Q06S	2.7469	0.0060**	1.7104	0.0872	-0.8469	0.003188	B	TBO
DR566Q12C	2.1045	0.0353*	0.7936	0.4274	1.2845	0.007333	B	TBO

Not. Referans grup: Erkek (N=78), Odak grup: Kadın (N=69); \*p<0.5, \*\*p<.01, \*\*\*p<.001

İşaretsiz alan testi sonucunda Raju'nun Z istatistiğinin manidarlığı değerlendirilmiş ve Tablo 14'te verilen 17 maddenin de referans grup (erkekler) lehine DMF içerdiği belirlenmiştir. İşaretili alan testi sonucunda Raju'nun Z istatistiğinin manidarlığı değerlendirildiğinde ise DR545Q04C, CR404Q07S ve DR404Q10BC maddelerinin odak grup (kadınlar) lehine; CR545Q06S, CR545Q07S, DR558Q04C ve CR543Q03S maddelerinin referans grup (erkekler) lehine DMF içerdiği belirlenmiştir.

Tablo 14'te p değerleri manidar çıkan maddelere ait NCDIF değerleri Oshima ve Wright (2015) tarafından önerilen etki büyüklüğü yorumlama ölçütleriyle karşılaştırılmış ve maddelerin hangi düzeyde DMF içerdiği belirlenmiştir. Buna göre beş maddenin orta düzeyde (B) ve 12 maddenin ihmal

edilebilir düzeyde (A) DMF içerdiği tespit edilmiştir. İşaretli ve işaretsiz alan indeksleri incelenmiş ve DMF biçimi belirlenmiştir. Buna göre bir maddede TB, diğer 16 maddede ise TBO-DMF tespit edilmiştir.

Cinsiyet değişkenine göre ele alınan yöntemlerde ortak olarak DMF belirlenen maddeler Tablo 15'te verilmiştir.

**Tablo 15.** Cinsiyet Değişkenine Göre Farklı Yöntemlerde Ortak Olarak DMF Belirlenen Maddeler ve DMF Düzeyleri

	MH	SIBTEST	LR	Raju'nun Alan Ölçüleri
DR558Q12C	C		B	
CR566Q05S	C		B	
CR545Q06S		C		B
CR558Q09S		C		A
CR566Q14S		C	B	

### İkinci Alt Probleme İlişkin Bulgular

PISA 2018 okuma becerileri alt testinde yer alan maddelerin ESCS değişkenine göre DMF içerip içermediği MH, SIBTEST, LR ve Raju'nun Alan Ölçüleri yöntemleriyle test edilmiştir.

MH yöntemine göre DMF içeren maddeler Tablo16'da görülmektedir.

**Tablo 16.** ESCS Değişkeni İçin MH Sonuçları

Maddeler	$\alpha_{MH}$	$\chi^2$	p	$\Delta_{MH}$	DMF Düzeyi	Hangi Grup Lehine
CR543Q10S	0.4536	-1.9635	0.0496*	1.8578	C	Odak grup
CR566Q04S	2.6879	2.4833	0.0130*	-2.3236	C	Referans grup

Not. Referans grup: Düşük (N=77), Odak grup: Yüksek (N=70); \*p<.05

Tablo 16'da p değerleri manidar olan maddelerin  $\Delta_{MH}$  değerleri incelenmiş olup Zieky (1993)'nin MH yöntemi için sınıfladığı  $\Delta_{MH}$  eşik değerleriyle karşılaştırılarak maddelerin hangi düzeyde DMF içerdiği belirlenmiştir. Buna göre iki maddenin yüksek düzeyde (C) DMF içerdiği tespit edilmiştir.

$\Delta_{MH}$  değerinin pozitif ya da negatif olmasına göre de DMF içeren maddenin hangi grubun lehine olduğu belirlenmiştir. CR543Q10S maddesinin odak grup lehine (yüksek ESCS) ve CR566Q04S maddesinin referans grup lehine (düşük ESCS) DMF içerdiği tespit edilmiştir.

SIBTEST yöntemine göre DMF içeren madde tespit edilmemiştir. LR yöntemine göre DMF içeren madde Tablo 17'de görülmektedir.

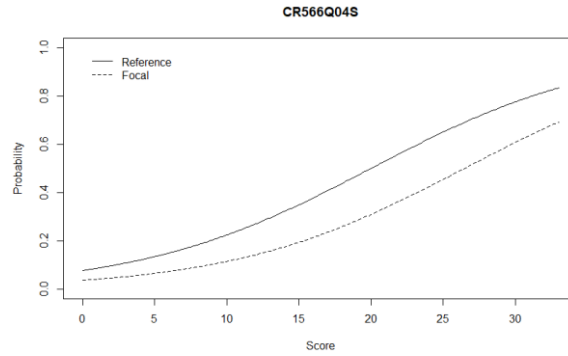
**Tablo 17.** ESCS Değişkeni İçin LR Sonuçları

Maddeler	TB DMF $R^2$	TBO DMF $R^2$	$\chi^2$	p	$R^2$	DMF Düzeyi Jodoin ve Gierl	DMF Biçimi
CR566Q04S	0.0423	0.0114	4.8156	0.0282*	0.0423	B	TB

Not. Referans grup: Düşük (N=77), Odak grup: Yüksek (N=70); \*p<.05

Tablo 17'de p değeri manidar olan maddelerin Nagelkerke  $R^2$  değerleri incelenmiş, Jodoin ve Gierl (2001)'in LR yöntemi için düzenlediği sınıflama sistemi kullanılarak maddelerin hangi düzeyde DMF içerdiği belirlenmiştir. Buna göre bir maddenin orta düzeyde (B) DMF içerdiği tespit edilmiştir.

DMF'nin biçimini belirlemek için TB- $R^2$  ve TBO- $R^2$  değerlerinin büyüklükleri karşılaştırılmıştır. Ardından madde düzeyinde grafiksel incelemeler yapılmıştır. Bu maddeye ait MKE Şekil 4'te verilmiştir.



Şekil 4. Madde Karakteristik Eğrisi (ESCS)

DMF tespit edilen maddenin grafiksel incelemesi sonucunda CR566Q04S maddesinin referans grup (düşük ESCS) lehine TB-DMF gösterdiği belirlenmiştir.

Raju'nun Alan Ölçüleri yöntemine göre DMF içeren maddeler Tablo 18'de görülmektedir.

Tablo 18. ESCS Değişkeni İçin Raju'nun Alan Ölçüleri Sonuçları

Maddeler	İşaretsiz Alan		İşaretli Alan		$\Delta_{MH}$	NCDIF = ( $\Delta_{MH} / K$ ) <sup>2</sup>	DMF Düzeyi	DMF Biçimi
	Z	p	Z	p				
DR545Q04C	-1.7045	0.0883	2.4640	0.0137*	1.3252	0.007805	B	TB
CR545Q07S	-2.2176	0.0266*	-1.2392	0.2153	0.3123	0.000433	A	TBO
CR404Q06S	-3.0876	0.0020**	2.1162	0.0343*	0.0118	0.000000	A	TBO
DR404Q10AC	-5.0506	0.0000***	3.3706	0.0008***	-0.1622	0.000117	A	TBO
DR404Q10BC	-5.2041	0.0000***	5.6692	0.0000***	-1.0609	0.005002	B	TB
DR558Q04C	-2.4803	0.0131*	-1.9545	0.0506	1.0110	0.004543	B	TBO
CR558Q06S	-2.4922	0.0127*	-1.2199	0.2225	-0.0830	0.000031	A	TBO
CR437Q01S	-2.5305	0.0114*	2.3427	0.0191*	-1.0418	0.004826	B	TBO
CR437Q06S	-2.4266	0.0152*	-0.0066	0.9947	0.2867	0.000365	A	TBO
CR543Q04S	-2.1331	0.0329*	-0.9490	0.3426	-1.3023	0.007538	B	TBO
CR543Q09S	-5.6812	0.0000***	4.2781	0.0000***	0.2399	0.000256	A	TBO
CR543Q13S	-3.5870	0.0003***	1.7668	0.0773	-1.3895	0.008581	C	TBO
DR543Q15C	-2.4356	0.0149*	2.5932	0.0095**	-0.6247	0.001734	A	TB
CR566Q04S	-2.2847	0.0223*	3.4435	0.0006***	-2.3236	0.023996	C	TB
CR566Q05S	-2.7794	0.0054**	2.4066	0.0161*	0.6934	0.002137	A	TBO
CR566Q06S	-2.7616	0.0058**	-0.3739	0.7085	0.4262	0.000807	A	TBO

Not. Referans grup: Düşük (N=77), Odak grup: Yüksek (N=70); \*p<.05

İşaretsiz alan testi sonucunda Raju'nun Z istatistiğinin manidarlığı değerlendirilmiş ve Tablo 18'de verilen maddelerden 15 maddenin odak grup (yüksek ESCS) lehine DMF içerdiği belirlenmiştir. İşaretli alan testi sonucunda Raju'nun Z istatistiğinin manidarlığı değerlendirildiğinde ise sekiz maddenin referans grup (düşük ESCS) lehine DMF içerdiği belirlenmiştir.

Tablo 18'de p değerleri manidar çıkan maddelere ait NCDIF değerleri Oshima ve Wright (2015) tarafından önerilen etki büyüklüğü yorumlama ölçütleriyle karşılaştırılmış ve maddelerin hangi düzeyde DMF içerdiği belirlenmiştir. Buna göre iki, beş ve dokuz maddenin sırasıyla yüksek düzeyde (C), orta düzeyde (B) ve ihmal edilebilir düzeyde (A) DMF içerdiği tespit edilmiştir. İşaretli ve işaretsiz alan indeksleri incelenmiş ve DMF biçimi belirlenmiştir. Buna göre dört maddede TB, diğer 12 maddede ise TBO-DMF tespit edilmiştir.

ESCS değişkenine göre ele alınan yöntemlerde ortak olarak DMF belirlenen maddeler Tablo 19'da verilmiştir.

**Tablo 19.** ESCS Değişkenine Göre Farklı Yöntemlerde Ortak Olarak DMF Belirlenen Maddeler ve DMF Düzeyleri

	MH	SIBTEST	LR	Raju'nun Alan Ölçüleri
CR566Q04S	C		B	C

### Üçüncü Alt Probleme İlişkin Bulgular

PISA 2018 okuma becerileri alt testinde yer alan maddelerin okulun bulunduğu yerleşim bölgesi değişkenine göre DMF içerip içermediği Genelleştirilmiş MH, Genelleştirilmiş LR ve Genelleştirilmiş Lord'un  $\chi^2$  yöntemleriyle test edilmiştir.

Genelleştirilmiş MH yöntemine göre DMF içeren maddeler Tablo20'de görülmektedir.

**Tablo 20.** Okulun Bulunduğu Yerleşim Bölgesi Değişkeni İçin Genelleştirilmiş MH Sonuçları

Maddeler	$\chi^2$	p
DR566Q12C	7.6484	0.0218*

Not. Referans grup: Büyük Şehir (N=62), Odak grup (4): Şehir (N=40), Odak grup (3): İlçe (N=45); \*p<.05

Tablo 20'de DR566Q12C maddesinin p değerinin manidar olduğu ve Genelleştirilmiş MH yöntemine göre DMF içerdiği belirlenmiştir. Genelleştirilmiş MH yönteminin birçok avantajı bulunmaktadır fakat DMF türü, etki büyüklüğü veya yönü ile ilgili ayrıntılı bilgi vermemektedir (Fidalgo ve Scalón, 2012).

Genelleştirilmiş LR yöntemine göre DMF içeren maddeler Tablo 21'de görülmektedir.

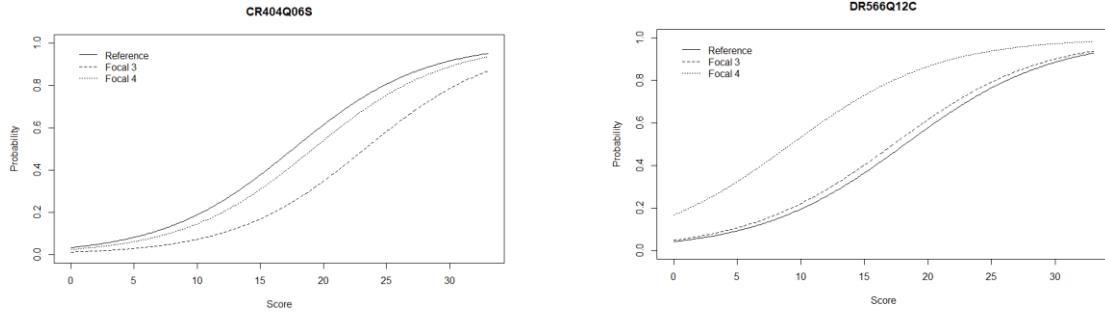
**Tablo 21.** Okulun Bulunduğu Yerleşim Bölgesi Değişkeni İçin Genelleştirilmiş LR Sonuçları

Maddeler	TB DMF $R^2$	TBO DMF $R^2$	$\chi^2$	p	$R^2$	DMF Düzeyi Jodoin ve Gierl	DMF Biçimi
CR404Q06S	0.0453	0.0114	6.0120	0.0495 *	0.0453	B	TB
DR566Q12C	0.0821	0.0188	8.6010	0.0136*	0.0821	C	TB

Not. Referans grup: Büyük Şehir (N=62), Odak grup (4): Şehir (N=40), Odak grup (3): İlçe (N=45); \*p<.05

Tablo 21'de p değeri manidar olan maddelerin Nagelkerke  $R^2$  değerleri incelenmiş, Jodoin ve Gierl (2001)'in LR yöntemi için düzenlediği sınıflama sistemi kullanılarak maddelerin hangi düzeyde DMF içerdiği belirlenmiştir. Buna göre bir maddenin orta düzeyde (B) ve bir maddenin de yüksek düzeyde (C) DMF içerdiği tespit edilmiştir.

DMF'nin biçimini belirlemek için  $TB-R^2$  ve  $TBO-R^2$  değerlerinin büyüklükleri karşılaştırılmıştır. Ardından maddelerin grafiksel incelemesi yapılmıştır. Bu maddelere ait MKE ise Şekil 5'te verilmiştir.



Şekil 5. Madde Karakteristik Eğrisi (Okulun Bulunduğu Yerleşim Bölgesi)

DMF tespit edilen maddelerin grafiksel incelemesi sonucunda CR404Q06S maddesinin ve DR566Q12C maddesinin sırasıyla referans grup (büyükşehir) ve odak grup – 4 (şehir) lehine TB-DMF içerdiği tespit edilmiştir.

Genelleştirilmiş Lord'un  $\chi^2$  yöntemine göre DMF içeren maddeler Tablo 22'de görülmektedir.

Tablo 22. Okulun Bulunduğu Yerleşim Bölgesi Değişkeni İçin Genelleştirilmiş Lord'un  $\chi^2$  Sonuçları

Maddeler	$\chi^2$	p
DR545Q04C	53.4426	0.0000 ***
CR545Q06S	78.9994	0.0000 ***
CR545Q07S	44.8603	0.0000 ***
CR424Q03S	15.4787	0.0038 **
CR424Q07S	160.8397	0.0000 ***
CR404Q03S	192.1537	0.0000 ***
CR404Q06S	24.7327	0.0001 ***
CR404Q07S	99.5895	0.0000 ***
DR404Q10AC	15.6739	0.0035 **
DR404Q10BC	48.8002	0.0000 ***
CR558Q02S	48.3320	0.0000 ***
DR558Q12C	17.3317	0.0017 **
DR558Q04C	76.3025	0.0000 ***
CR558Q06S	66.0696	0.0000 ***
CR558Q09S	41.3370	0.0000 ***
CR437Q06S	15.6896	0.0035 **
CR543Q01S	23.2275	0.0001 ***
CR543Q03S	105.8070	0.0000 ***
CR543Q04S	40.0946	0.0000 ***
CR543Q09S	24.0898	0.0001 ***
CR543Q10S	24.4286	0.0001 ***
CR543Q13S	17.4853	0.0016 **
DR543Q15C	209.3311	0.0000 ***
DR566Q03C	74.9375	0.0000 ***
CR566Q05S	16.0128	0.0030 **
CR566Q14S	30.3568	0.0000 ***
CR566Q06S	37.5355	0.0000 ***
DR566Q12C	60.4916	0.0000 ***

Not. Referans grup: Büyük Şehir (N=62), Odak grup (4): Şehir (N=40), Odak grup (3): İlçe (N=45); \*\*p<.01, \*\*\*p<.001



Tablo 22’de 28 maddenin p değerinin manidar olduğu ve Genelleştirilmiş Lord’un  $\chi^2$  yöntemine göre DMF içerdiği belirlenmiştir.

Okulun bulunduğu yerleşim bölgesi değişkenine göre ele alınan yöntemlerde ortak olarak DMF belirlenen maddeler Tablo 23’te verilmiştir.

**Tablo 23.** Okulun Bulunduğu Yerleşim Bölgesi Değişkenine Göre Farklı Yöntemlerde Ortak Olarak DMF Belirlenen Maddeler ve DMF Düzeyleri

	Genelleştirilmiş MH	Genelleştirilmiş LR	Genelleştirilmiş Lord’un $\chi^2$
DR566Q12C	x	C	x
CR404Q06S		B	x

### Sonuç, Tartışma ve Öneriler

Bu çalışmada kapsamında PISA 2018 okuma becerileri alt testinde yer alan maddelerin cinsiyet, ESCS ve okulun bulunduğu yerleşim bölgesi değişkenlerine göre DMF içerip içermediğinin farklı yöntemler ile analiz edilmesi ve sonuçların karşılaştırılması amaçlanmıştır. Birinci alt problemin sonucunda cinsiyet değişkenine göre MH, SIBTEST ve LR yöntemlerinde sırasıyla iki maddenin yüksek düzeyde (C), dört maddenin yüksek düzeyde (C) ve üç maddenin orta düzeyde (B) DMF gösterdiğine ulaşılmıştır. Raju’nun Alan Ölçüleri yönteminde ise beş maddenin orta düzeyde (B) ve 12 maddenin ihmal edilebilir düzeyde (A) DMF içerdiği sonucuna ulaşılmıştır. Ek olarak bazı maddelerin farklı yöntemlerden elde edilen sonuçlarda ortak şekilde DMF içerdiği görülmüştür: MH ve LR yöntemlerinde iki, LR ve SIBTEST yöntemlerinde bir, SIBTEST ve Raju’nun Alan Ölçüleri yöntemlerinde ise iki maddenin ortak olarak DMF içerdiği belirlenmiştir. Espinoza (2019)’nın PISA 2015 okuma becerisi alt testinde yer alan maddeler için cinsiyet bağlamında yürüttüğü DMF çalışmasında da çoğu maddenin –A düzeyinde de olsa– DMF içerdiği bulunmuştur. Ancak PISA gibi oldukça önemli bir geniş ölçekli test uygulamasında yer alan maddelerin DMF içeriyor olması şaşırtıcı bir bulgu değildir. Yalnızca PISA değil, uluslararası düzeyde uygulanan testler için cinsiyet üzerinden yürütülen farklı DMF çalışmalarında da benzer sonuçlara ulaşıldığı görülmektedir (Lan, 2014; Le, 2009; Lyons-Thomas, Sandilands ve Ercikan, 2014; Kıbrıslıoğlu Uysal ve Atalay Kabasakal, 2017).

İkinci alt problemde, ESCS değişkenine göre, MH ve LR yöntemlerinde sırasıyla iki maddenin yüksek düzeyde (C) ve bir maddenin orta düzeyde (B) DMF içerdiğine ulaşılmıştır. Raju’nun Alan Ölçüleri yönteminde ise sırasıyla iki, beş ve dokuz maddenin yüksek düzeyde (C), orta düzeyde (B) ve ihmal edilebilir düzeyde (A) DMF içerdiği tespit edilirken, SIBTEST yönteminde DMF içeren madde bulunmamıştır. İlgili maddeler incelendiğinde, “CR566Q04S” kodlu maddenin her üç yöntemde de ortak olarak DMF içerdiği belirlenmiştir. Bu sonuçlar Espinoza (2019)’nın PISA 2015’te; Chen ve Jiao (2014)’nın PISA 2009’da yer alan okuma becerileri testindeki maddeler için yürüttüğü DMF çalışmaları ile benzerdir.

Üçüncü alt problemde, okulun bulunduğu yerleşim bölgesi değişkenine göre, Genelleştirilmiş LR yönteminde bir maddenin orta düzeyde (B) ve bir maddenin de yüksek düzeyde (C) DMF içerdiği tespit edilmiştir. Genelleştirilmiş MH ve Genelleştirilmiş Lord’un  $\chi^2$  yöntemlerinde ise sırasıyla bir maddenin ve 28 maddenin DMF içerdiğine ulaşılmıştır. İlgili maddeler incelendiğinde bir maddenin Genelleştirilmiş LR ve Genelleştirilmiş Lord’un  $\chi^2$  yöntemlerinde, “DR566Q12C” kodlu maddenin ise her üç yöntemde de ortak olarak DMF içerdiği belirlenmiştir. Literatür incelendiğinde, odak grup sayısının ikiden fazla olduğu Genelleştirilmiş MH ve Genelleştirilmiş LR yöntemlerinde benzer sonuçların elde edildiği görülmüştür. Bu çalışmaya benzer olarak, Uyar ve Uyanık (2016), Genelleştirilmiş MH ve Genelleştirilmiş LR yöntemlerini kullanarak DMF incelemesi yaptıkları çalışmalarında, kullanılan yöntemlerde yaklaşık olarak aynı sonuçlara ulaşmıştır.

Bu sonuçlar genel olarak değerlendirildiğinde, KTK ve MTK’ya dayalı yöntemlerin kendi içerisinde uyumlu olduğu ve KTK’ya dayalı yöntemlerin MTK’ya dayalı yöntemlere göre çok daha az sayıda maddeyi DMF’li olarak tespit ettiği sonuçlarına ulaşılabilir. Bu farklılığın MTK’ya dayalı yöntemlerin daha hassas olmasından ve en küçük farklılığı bile DMF olarak tespit etmesinden kaynaklandığı düşünülmektedir. Mazor, Clauser ve Hambleton (1994), MTK tabanlı yöntemlerin tek

biçimli olmayan DMF'ye duyarlı olduğunu fakat bu yöntemlerin örneklem boyutu ve karmaşık hesaplama gereksinimleri olduğunu belirtmiştir. Benzer şekilde, Atalay, Gök, Kelecioğlu ve Arsan (2012) çalışmalarında örtük puan yöntemlerinin gözlenen puan yöntemlerine göre DMF'li maddeleri belirlemede daha duyarlı ve etkili olduğu sonucuna ulaşmıştır. Pektaş (2018) çalışmasında, DMF özelliği gösteren maddelerin belirlenmesinde, KTK ve MTK'ya dayalı yöntemler arasında farklılık olduğu; MTK'ya dayalı yöntemlerde daha fazla DMF'li madde kestiriminin gerçekleştirildiğini belirtmiştir. DMF belirlenirken kullanılan yöntemlerde farklı sonuçlar elde edilmesinin en temel sebebi, bu yöntemlerin DMF analizlerinde farklı istatistiksel yollar ve farklı aşamalar kullanmasıdır. Bu analizlerden elde edilen değerlere göre DMF kararı verilmekte fakat elde edilen sonuçlara farklı matematiksel yöntemlerle ulaşılmaktadır (Uzun ve Gelbal, 2017). Bu çalışmada da görüldüğü gibi, KTK ve MTK'ya dayalı yöntemler kendi içerisinde tutarlık gösterse de DMF düzeylerinde farklılık görülebilmektedir.

Cinsiyet ve ESCS değişkenlerinde birer madde hem MH hem de LR yöntemine göre DMF'li olarak tespit edilmiştir. MH ve LR yöntemlerinin diğer yöntemlere göre daha uyumlu sonuçlar çıkarması Gök ve diğerleri (2010) ve Yurdugül (2003)'ün çalışmalarının sonuçlarıyla tutarlıdır. Fakat cinsiyet ve ESCS değişkenlerinde ortak olarak DMF tespit edilen bu maddelerin DMF düzeyleri farklıdır. Bu maddeler MH tekniğinde yüksek düzeyde DMF içerirken LR tekniğinde orta düzeyde DMF içermektedir. DMF düzeylerindeki bu farklılığın, farklı yöntemlerin DMF düzeylerini sınıflandırırken kullandıkları değer aralıklarının farklı olmasından kaynaklanabileceği düşünülmektedir. İlgili yöntemlerin her ikisi de toplam test puanları üzerinden analiz yapıyor olmasına ve benzer bileşenleri bulunmasına rağmen kategorilendirmede kullandıkları ölçütün farklılık oluşturduğu düşünülmektedir.

Elde edilen bulgular incelendiğinde, yöntemlerin TB ve TBO-DMF'yi tespit etmede farklılaştığı belirlenmiştir. MH yöntemi TB-DMF'yi tespit etmek için tasarlanırken LR yöntemi TBO-DMF'yi tespit etmek için tasarlanmıştır (Swaminathan ve Rogers, 1990). Swaminathan ve Rogers (1990) ve Rogers ve Swaminathan (1993) yaptıkları çalışmada, LR yönteminin TBO-DMF'yi tespit etmede MH yönteminden daha güçlü olduğuna; TB-DMF'yi tespit etmede ise MH yöntemi ile aynı güce sahip olduğuna ulaşmıştır. Benzer şekilde, Narayanan ve Swaminathan (1996) çalışmalarında, TBO-DMF'yi tespit etmede, SIBTEST ve LR yöntemlerinin MH yönteminden daha güçlü olduğuna ulaşmıştır. Clauser ve Mazor (1998) ise SIBTEST yöntemiyle elde edilen sonuçların MH ile karşılaştırılabilir olduğunu ancak SIBTEST'in TBO-DMF'yi tespit ettiğini, bu sebeple de LR yöntemi ile elde edilen bulguların kontrol edilebilmesinde kullanıldığını ifade etmiştir.

Literatür incelendiğinde, KTK ve MTK'ya dayalı DMF belirleme yöntemlerinin karşılaştırıldığı birçok çalışma görülmüştür. KTK kapsamında yapılan madde yanlılığı çalışmaları, kuramın bazı dezavantajları göz önünde bulundurularak, eleştirilmiştir (Shepard, Camilli ve Williams, 1985). Bu eleştiriler, KTK'ya dayalı yöntemlerde madde parametrelerinin bir gruptan diğerine değişmesi, sabit olmaması ya da örnekleme bağlı olması sebebiyle yapılan DMF analizlerinin yanlış yorumlara yol açabileceği şeklindedir. MTK'ya dayalı yöntemlerin KTK'ya dayalı yöntemlere göre daha üstün yönlerinin bulunduğu ileri sürülmektedir (Lord, Novick ve Birnbaum, 1968; Hambleton ve Swaminathan 1985). MTK'ya dayalı yöntemlerde madde parametrelerinin bir gruptan diğerine değişmemesi, yani sabit olması, bu sebeple de grupları karşılaştırmada maddenin yanlı olup olmadığına karar vermede daha doğru sonuçlar verdiği düşüncesi kesin kanıtlar olmamasına rağmen yaygınlaşmıştır. Bu görüşe uygun olarak Clauser ve Mazor (1998) çalışmasında MTK'ya dayalı yöntemlerin en çok savunulan yöntemler olduğunu belirtmiştir. Fakat MTK'ya dayalı yöntemlerdeki ortak problem, kuramın varsayımlarının karşılanmasının ve model veri uyumunun sağlanmasının kestirimlerin güvenilirliğini etkileyebilmesidir. Bu çalışmada da örneklem büyüklüğü MTK varsayımları açısından bir sınırlılık olarak belirlenmiştir. Ancak, Belzak (2020), örneklem büyüklüğünün küçük olduğu durumlarda, DMF tespit etmenin daha az karmaşık modellerde daha doğru sonuçlar verdiğini ortaya koymuştur.

Sonuçları genel olarak deđerlendirmek gerekirse, PISA 2018 okuma becerileri alt testinde yer alan ve 1-0 şeklinde puanlanan maddelerin madde yanlılıđı aısından tartıřılabilir olduđu sylenbilir. DMF ieren madde sayısının fazla olması bu testlerle yapılan karřılařtırmalarda dođru yorumlar yapılmasını engelleyecektir. Dolayısıyla, PISA testlerinde maddelerin bazı grupların lehine avantaj sađlıyor grnmesi bu testlerle yapılan yorumların geerliđini ve gvenirliđini tartıřılır hale getirmektedir. Fakat burada dikkat edilmesi gereken nokta, DMF ieren her bir maddenin yanlılık gstermeyebileceđidir. nk bu farklılık maddenin gerek etkisinden de kaynaklanıyor olabilir. Bu sebeple, DMF ieren maddeler uzmanlar tarafından ayrıntılı olarak incelenmeli ve gruplarda nasıl alıřtıđı tespit edilmelidir. Ek olarak, diđer lkelerde de bu alıřmalar yrtlp benzer sonulara ulařıldıđında genellemeler yapılabilen iin, PISA gibi uluslararası test uygulamalarında DMF alıřmalarının dzenli bir şekilde yapılması nerilebilir.

## Kaynakça

- Acar, T. (2011). Sample size in differential item functioning: An application of hierarchical linear modeling. *Kuram ve Uygulamada Eğitim Bilimleri*, 11(1), 284-288.
- Akbaşı, S., Şahin, M. ve Yaykırın, Z. (2016). The effect of reading comprehension on the performance in science and mathematics. *Journal of Education and Practice*, 7(16), 108-121.
- Atalay, K., Gök, B., Kelecioğlu, H. ve Arsan, N. (2012). Değişen Madde Fonksiyonunun belirlenmesinde kullanılan farklı yöntemlerin karşılaştırılması: Bir simülasyon çalışması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 43(43), 270-281.
- Ateşok Deveci, N. (2008). *Üniversitelerarası kurul yabancı dil sınavının madde yanlılığı bakımından incelenmesi* (Yayımlanmamış doktora tezi). Ankara Üniversitesi, Ankara.
- Bakan Kalaycıoğlu, D. (2008). *Öğrenci Seçme Sınavı'nın madde yanlılığı açısından incelenmesi* (Yayımlanmamış doktora tezi). Hacettepe Üniversitesi Ankara.
- Bakan Kalaycıoğlu, D. ve Kelecioğlu, H. (2011). *Öğrenci Seçme Sınavının madde yanlılığı açısından incelenmesi*. *Eğitim ve Bilim*, 36(161), 3-13.
- Belzak, W. C. (2020). Testing differential item functioning in small samples. *Multivariate Behavioral Research*, 55(5), 722-747.
- Berberoğlu, G. (1995). Differential item functioning (DIF) analysis of computation, word problem and geometry questions across gender and SES groups. *Studies in Educational Evaluation*, 21(4), 439-455.
- Birjandi, P. ve Amini, M. (2007). Differential item functioning (test bias) analysis paradigm across manifest and latent examinee groups (on the construct validity of IELTS). *Journal of Human Sciences*, 8(2), 1-20.
- Büyüköztürk, Ş., Çakmak, E. K., Akgün, Ö. E., Karadeniz, Ş. ve Demirel, F. (2009). *Bilimsel araştırma yöntemleri*. Ankara: Pegem Akademi.
- Camilli, G. ve Shepard, L. A. (1994). *Methods for identifying biased test items*. Newbury Park, CA: Sage.
- Chen, Y. F. ve Jiao, H. (2014). Exploring the utility of background and cognitive variables in explaining latent differential item functioning: An example of the PISA 2009 reading assessment. *Educational Assessment*, 19(2), 77-96.
- Clauser, B. E. ve Mazor, K. M. (1998). Using statistical procedures to identify differentially functioning test items. *Educational Measurement: Issues and Practice*, 17(1), 31-44.
- Çelik, M. ve Özkan, Y. Ö. (2020). Analysis of differential item functioning of PISA 2015 mathematics subtest subject to gender and statistical regions. *Journal of Measurement and Evaluation in Education and Psychology*, 11(3), 283-301.
- Çet, S. (2006). *PISA 2003 matematik maddeleri kullanılarak yanlı çalışan maddelerin tespitinde çok boyutlu eşleştirme analizi* (Yayımlanmamış doktora tezi). Orta Doğu Teknik Üniversitesi, Ankara.
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. New York: The Guilford Press.
- DeMars, C. (2016). *Madde tepki kuramı* (H. Kelecioğlu, Çev. ve Ed.). Ankara: Nobel Akademi.
- Ding, H. ve Homer, M. (2020). Interpreting mathematics performance in PISA: Taking account of reading performance. *International Journal of Educational Research*, 102, 101566.
- Ercikan, K. ve Kim, K. (2005). Examining the construct comparability of the English and French versions of TIMSS. *International Journal of Testing*, 5(1), 23-35.
- Erdem, E. (2016). Relationship between mathematical reasoning and reading comprehension: The case of the 8th grade. *Necatibey Faculty of Education Electronic Journal of Science & Mathematics Education*, 10(1), 393-414.
- Espinoza, J. C. (2019). *Differential item functioning analysis of PISA 2015 reading items: Singapore, Australia and USA* (Yayımlanmamış doktora tezi). Hacettepe Üniversitesi, Ankara.

- Fidalgo, Á. M. ve Scalon, J. D. (2012). Using Mantel-Haenszel methods for detecting differential item functioning. *Psicologia, Reflexão e Crítica*, 25(1), 60.
- Fuentes, P. (1998). Reading comprehension in mathematics. *The Clearing House*, 72(2), 81-88.
- Geske, A. ve Ozola, A. (2008). Factors influencing reading literacy at the primary school level. *Problems of Education in the 21st Century*, 6, 71-77.
- Gök, B., Kabasakal, K. A. ve Kelecioğlu, H. (2014). PISA 2009 öğrenci anketi tutum maddelerinin kültüre göre değişen madde fonksiyonu açısından incelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 5(1), 72-87.
- Gök, B., Kelecioğlu, H. ve Doğan, N. (2010). Değişen madde fonksiyonu belirlemede Mantel -Haenzel ve Lojistik Regresyon tekniklerinin karşılaştırılması. *Eğitim ve Bilim*, 35(156), 3-16.
- Grimm, K. J. (2008). Longitudinal associations between reading and mathematics achievement. *Developmental Neuropsychology*, 33(3), 410-426. doi:10.1080/87565640801982486
- Grisay, A. ve Monseur, C. (2007). Measuring the equivalence of item difficulty in the various versions of an international test. *Studies in Educational Evaluation*, 33(1), 69-86.
- Gür, E. (2019). *PISA 2015 uygulamasındaki maddelerin kültüre göre değişen madde fonksiyonu açısından incelenmesi* (Yayımlanmamış yüksek lisans tezi). Hacettepe Üniversitesi, Ankara.
- Hambleton, R. K. ve Swaminthan, H. (1985). *Item response theory: Principles and applications*. Boston, MA: Kluwer-Nijhoff.
- Hambleton, R. K., Swaminathan, H. ve Rogers, H. J. (1991). *Fundamentals of item response theory* (2. cilt). Thousand Oaks, CA: Sage.
- Hamilton, L. S. ve Snow, R. E. (1998). *Exploring differential item functioning on science achievement tests* (CSE Teknik Rapor No. 483). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing, University of California.
- Jodoin, M. G. ve Gierl, M. J. (2001). Evaluating type I error and power rates using an effect size measure with the logistic regression procedure for DIF detection. *Applied Measurement in Education*, 14(4), 329-349.
- Kabasakal, K. A. ve Kelecioğlu, H. (2012). Evaluation of attitude items in PISA 2006 student questionnaire in terms of differential item functioning. *Ankara University Journal of Faculty of Educational Sciences (JFES)*, 45(2), 77-96.
- Karakaya, İ. ve Kutlu, Ö. (2012). Seviye belirleme sınavındaki Türkçe alt testlerinin madde yanlılığının incelenmesi. *Eğitim ve Bilim*, 37(165), 348-362.
- Kelecioğlu, H., Karabay, B. ve Karabay, E. (2014). Seviye Belirleme Sınavı'nın madde yanlılığı açısından incelenmesi. *İlköğretim Online*, 13(3), 934-953.
- Kıbrıslıoğlu Uysal, N. ve Atalay Kabasakal, K. (2017). The effect of background variables on gender related differential item functioning. *Journal of Measurement and Evaluation in Education and Psychology*, 8(4), 373-390.
- Kim, S. H., Cohen, A. S. ve Park, T. H. (1995). Detection of differential item functioning in multiple groups. *Journal of Educational Measurement*, 32(3), 261-276.
- Kristjansson, E., Aylesworth, R., Mcdowell, I. ve Zumbo, B. D. (2005). A comparison of four methods for detecting differential item functioning in ordered response items. *Educational and Psychological Measurement*, 65(6), 935-953.
- Lan, M. C. (2014). *Exploring gender differential item functioning (DIF) in eight grade mathematics items for the United States and Taiwan* (Doktora tezi). Washington Üniversitesi, Washington.
- Le, L. T. (2006, Nisan). *Analysis of differential item functioning*. Meeting of the American Educational Research Association'da sunulan bildiri, San Francisco CA.
- Le, L. T. (2009). Investigating gender differential item functioning across countries and test languages for PISA science items. *International Journal of Testing*, 9(2), 122-133.

- Lerikkanen, M. K., Rasku-Puttonen, H., Aunola, K. ve Nurmi, J. E. (2005). Mathematical performance predicts progress in reading comprehension among 7-year olds. *European Journal of Psychology of Education*, 21(2), 121-137.
- Little R. J. A ve Rubin D. R. (2002). *Statistical analysis with missing data* (2. bs.). New York: Wiley.
- Lord, F. M., Novick, M. R. ve Birnbaum, A. (1968). *Statistical theories of mental test scores*. New York: Addison-Wesley Publishing Company.
- Lyons-Thomas, J., Sandilands, D. D. ve Ercikan, K. (2014). Gender differential item functioning in mathematics in four international jurisdictions. *Education and Science*, 39(172), 20-32.
- Magis, D., Raîche, G., Béland, S. ve Gérard, P. (2011). A generalized logistic regression procedure to detect differential item functioning among multiple groups. *International Journal of Testing*, 11(4), 365-386.
- Mazor, K. M., Clauser, B. E. ve Hambleton, R. K. (1994). Identification of nonuniform differential item functioning using a variation of the Mantel-Haenszel procedure. *Educational and Psychological Measurement*, 54(2), 284-291.
- Milli Eğitim Bakanlığı. (2019). *PISA 2018 Türkiye ön raporu*. Ankara: MEB.
- Narayanan, P. ve Swaminathan, H. (1994). Performance of the Mantel-Haenszel and simultaneous item bias procedures for detecting differential item functioning. *Applied Psychological Measurement*, 18(4), 315-328.
- Narayanan, P. ve Swaminathan, H. (1996). Identification of items that non-uniform DIF. *Applied Psychological Measurement*, 20(3), 257-274.
- OECD. (2019). *PISA 2018 results volume I: What students know and can do*. Paris: OECD Publishing.
- Oshima, T. C. ve Wright, K. D. (2015). An effect size measure for Raju's differential functioning for items and tests. *Educational and Psychological Measurement*, 75(2), 338-358.
- Osterholm, M. (2005). Characterizing reading comprehension of mathematical texts. *Educational Studies in Mathematics*, 63, 325- 346.
- Osterlind, S. J. (1983). *Test item bias*. Thousand Oaks, CA: Sage.
- Öğretmen, T. ve Doğan, N. (2004). OKÖSYS Matematik alt testine ait maddelerin yanlılık analizi. *İnönü Üniversitesi Eğitim Fakültesi Dergisi*, 5(8), 61-76.
- Pektaş, S. (2018). *Değişen madde fonksiyonu belirleme yöntemlerinin test parametreleri kestirimlerine, karar çalışmalarına, g ve phi katsayılarına etkisi* (Yayımlanmamış doktora tezi). Gazi Üniversitesi, Ankara.
- Rindermann, H. ve Baumeister, A. E. E. (2015). Parents' SES vs. parental educational behavior and children's development: A reanalysis of the Hart and Risley study. *Learning and Individual Differences*, 37, 133-138.
- Rogers, H. J. ve Swaminathan, H. (1993). A comparison of logistic regression and Mantel-Haenszel procedures for detecting differential item functioning. *Applied Psychological Measurement*, 17(2), 105-116.
- Roussos, L. ve Stout, W. (1996). A multidimensionality-based DIF analysis paradigm. *Applied Psychological Measurement*, 20(4), 355-371.
- Shealy, R. ve Stout, W. (1993). A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DTF as well as item bias/DIF. *Psychometrika*, 58(2), 159-194.
- Shepard, L. A., Camilli, G. ve Williams, D. M. (1985). Validity of approximation techniques for detecting item bias. *Journal of Educational Measurement*, 22(2), 77-105.
- Sırgancı, G. (2012). *PISA 2006 öğrenci anketi madde yanlılığının sıralı lojistik regresyon ve poly-SIBTEST yöntemleri ile test edilmesi* (Yayımlanmamış yüksek lisans tezi). Abant İzzet Baysal Üniversitesi, Bolu.

- Swaminathan, H. ve Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27(4), 361-370.
- Şenferah, S. (2015). *2010 Seviye belirleme sınavı matematik alt testi için değişen madde fonksiyonlarının ve madde yanlılığının incelenmesi* (Yayımlanmamış yüksek lisans tezi). Gazi Üniversitesi, Ankara.
- Tabachnick, B. G. ve Fidell, L. S. (2015). *Çok değişkenli istatistiklerin kullanımı* (M. Baloğlu, Çev. ve Ed.). Ankara: Nobel Akademi.
- Thissen, D., Steinberg, L. ve Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. P. W. Holland ve H. Wainer (Ed.), *Differential item functioning* içinde (s. 35-113). Hillsdale, NJ: Lawrence Erlbaum.
- Ulutaş, S. (2012). *PISA 2006 fen okuryazarlığı testindeki maddelerin yanlılık bakımından araştırılması* (Yayımlanmamış doktora tezi). Ankara Üniversitesi, Ankara.
- Uyar, Ş. ve Uyanık, G. K. (2016). PISA 2012 Bilişsel maddelerinin kültüre göre değişen madde fonksiyonu bakımından incelenmesi. *Eğitim ve Öğretim Araştırmaları Dergisi*, 5(3), 230-240.
- Uzun, N. B. ve Gelbal, S. (2017). PISA fen başarı testinin madde yanlılığının kültür ve dil açısından incelenmesi. *Kastamonu Eğitim Dergisi*, 25(6), 2427-2446.
- Walzebug, A. (2014). Is there a language-based social disadvantage in solving mathematical items?. *Learning, Culture and Social Interaction*, 3(2), 159-169.
- Wang, W. C. ve Su, Y. H. (2004). Effects of average signed area between two item characteristic curves and test purification procedures on the DIF detection via the Mantel-Haenszel method. *Applied Measurement in Education*, 17(2), 113-144.
- Wright, B. D. ve Masters, G. N. (1982). *Rating scale analysis*. Kaliforniya: MESA Press.
- Yıldırım, H. (2015). *2012 yılı seviye belirleme sınavı matematik alt testinin madde yanlılığı açısından incelenmesi* (Yayımlanmamış yüksek lisans tezi). Gazi Üniversitesi, Ankara.
- Yurdugül, H. (2003). *Ortaöğretim kurumları seçme ve yerleştirme sınavının madde yanlılığı açısından incelenmesi* (Yayımlanmamış doktora tezi). Hacettepe Üniversitesi, Ankara.
- Yurdugül, H. ve Aşkar, P. (2004). Ortaöğretim Kurumları Öğrenci Seçme Ve Yerleştirme Sınavı'nın, öğrencilerin yerleşim yerlerine göre, diferansiyel madde fonksiyonu açısından incelenmesi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 27(27), 268-275.
- Zenisky, A. L., Hambleton, R. K. ve Robin, F. (2004). DIF detection and interpretation in large-scale science assessments: Informing item writing practices. *Educational Assessment*, 9(1-2), 61-78.
- Zieky, M. (1993). Practical questions in the use of DIF statistics in test development. P. W. Holland ve H. Wainer (Ed.), *Differential item functioning* içinde (ss. 337-347). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores*. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.
- Zumbo, B. D. ve Thomas, D. R. (1996). *A measure of DIF effect size using logistic regression procedures*. National Board of Medical Examiners'da sunulan bildiri, Philadelphia.