

Öğrenci Özelliklerinin Cinsiyete Dayalı Değişen Madde Fonksiyonuna Etkisi*

The Effect of Background Variables on Gender Related Differential Item Functioning

Nermin KIBRISLIOĞLU UYSAL**

Kübra ATALAY KABASAKAL***

Öz

Araştırmada, sosyoekonomik düzey ve okuma becerisinin cinsiyete dayalı DMF'nin ortaya çıkmasına etkisi amaçlanmıştır. Bu kapsamda PISA 2015 fen uygulamasında yer alan maddelerin ele alınan dokuz ülkede cinsiyete göre değişen madde fonksiyonu (DMF) gösterip göstermediğinin belirlenmiştir. Araştırma kapsamında bilgisayar tabanlı uygulamada yer alan bir fen madde kümesi ele alınmıştır. Araştırmaya dahil edilen ülkeler bilgisayar tabanlı uygulamaya katılan ülkeler arasında başarı sıralamalarına göre seçilmiştir. Analizlerde çoklu göstergeler çoklu nedenler modeli (MIMIC) yöntemi kullanılmıştır. MIMIC ile DMF analizleri belirli gruplarda maddelerin örtük özelliği eşit şekilde ölçüp ölçmediğini belirlemek amacıyla, tam ve sınırlandırılmış modellerin uyumlarının karşılaştırılması içerir. Analizler iki aşamada gerçekleştirilmiştir. İlk aşamada maddelerin cinsiyet grupları arasında DMF gösterip göstermediği incelenmiştir. Daha sonra sosyoekonomik düzey ve okuma becerisi değişkenleri sırasıyla modele eklenerek söz konusu değişkenlerin cinsiyetten kaynaklı DMF'ye etkisi incelenmiştir. Araştırmanın bulgularına göre seçilen ülkelerin tamamında cinsiyetle ilişkili DMF'li maddeler yer almakta ve DMF'li madde sayıları 2 ile 6 arasında değişmektedir. Ülkelerin dördünde modele eklenen değişkenler cinsiyete ilişkin DMF'li madde sayısını manidar şekilde etkilememiştir. Ancak diğer dört ülkede söz konusu değişkenlerin modele eklenmesi DMF'li madde sayısını azaltmıştır. Cinsiyete dayalı DMF'li madde sayısını azaltan değişkenler her bir ülke kapsamında ayrı ayrı tartışılmıştır.

Anahtar Kelimeler: Sosyoekonomik düzey, okuma başarısı, değişen madde fonksiyonu, MIMIC

Abstract

In this study, the effect of socioeconomic status and reading ability, on the presence of gender-related DIF were examined. For this purpose, presence of differential item functioning (DIF) between gender groups in PISA 2015 science items in nine selected countries were detected. One cluster of science items from computer-based assessment (CBA) was taken into consideration. The countries were selected among the ones that implemented CBA, on the basis of their rank in science achievement. Multiple Indicator Multiple Causes method (MIMIC) was used for DIF analyses. DIF analysis in the MIMIC involves fit comparisons of both full and reduced models to determine if the items can measure the latent trait equally among the specified groups. The MIMIC analysis was conducted in two steps. First, the items were tested for exhibiting DIF between gender groups. Then the socioeconomic status and the reading ability were added to the model to test gender-related DIF items and their effects, respectively. According to the results of the study, gender-related DIF appeared in all of the selected countries with between two and six items. In four of the countries, none of the selected variables significantly affected the presence of gender-related DIF. Instead, in the remaining countries, the number of gender-related

* Bu çalışma "12th International Conference on Social Sciences" kongresinde sözlü bildiri olarak sunulmuştur.

** Arş. Gör., Hacettepe Üniversitesi, Eğitim Bilimleri Bölümü, Ankara-Türkiye, e-posta: nkibrislioglu@hacettepe.edu.tr, ORCID ID: orcid.org/0000-0002-9592-469X

*** Yrd. Doç. Dr., Hacettepe Üniversitesi, Eğitim Bilimleri Bölümü, Ankara-Türkiye, e-posta: katalay@hacettepe.edu.tr, ORCID ID: orcid.org/0000-0002-3580-5568

DIF items was decreased by adding selected variables to the model. The effects of variables which reduced the number of gender-related DIF items were discussed within each country.

Keywords: Socioeconomic status, reading ability, differential item functioning, MIMIC

GİRİŞ

Uluslararası Öğrenci Değerlendirme Programı (PISA), Ekonomik İşbirliği ve Kalkınma Örgütü (OECD) tarafından 2000 yılında uygulanmaya başlayan ve üç yıl aralıklarla tekrarlanan okuma, matematik ve fen bilimleri temel alanlarındaki yeterliği ölçmeyi amaçlayan büyük ölçekli bir araştırmadır. PISA, 15 yaş grubu öğrencilerinin bu üç temel alanda okul programlarından ziyade, gerçek yaşamda karşılarına çıkabilecek sorunlardaki bilgi ve becerilerini kullanabilme kapasitelerini değerlendirir. PISA sonuçları ülkelerin eğitim politikalarını önemli düzeyde etkilemesi, ülkelerin kendi başarılarını diğer ülkelerle karşılaştırmalı olarak değerlendirmesine olanak sağlaması bakımından oldukça önemlidir. Bu karşılaştırmalar yapılırken farklı kültürlerden ve dillerden gelen bireyler, bilişsel ya da duyuşsal özellikleri bakımından sıralanır. Bu nedenlerle PISA sonuçlarının farklı ülkeler arasında karşılaştırılabilirliği ile ilgili endişeler son yıllarda artmaya başlamıştır.

Ölçme araçlarında istenilen, aracın psikometrik özelliklerinin tüm cevaplayıcılar için aynı olmasıdır (Woods, 2009). Bir ölçme aracındaki maddelerin psikometrik özelliklerinin aynı yeteneğe sahip farklı gruplar için aynı olmaması durumu yanlılık olarak tanımlanır. Psikometride madde düzeyinde yanlılığın test edildiği istatistiksel analizlerden biri değişen madde fonksiyonudur (DMF). DMF, aynı yetenek düzeyine sahip farklı gruptaki bireylerin maddeyi doğru cevaplama olasılıklarının farklılaşmasıdır. Madde Tepki Kuramı (MTK) çerçevesinde, DMF gösteren bir madde farklı gruplar için farklı kategori cevap fonksiyonlarına sahip olacaktır. Diğer bir ifadeyle, örtük değişken (θ) üzerinde eşleşen iki farklı grubun bireylerinin, aynı maddeye aynı yanıtı verme olasılıkları aynı olmayacaktır. Gruplar cinsiyet, etnik köken ya da deneysel duruma göre tanımlanabilir. Genellikle çoğunluğu oluşturan ya da testin orijinalinin geliştirildiği grup referans grup, diğer grup odak grup olarak isimlendirilir (Zumbo, 1999). Referans ve odak gruba dayalı olarak uygulanan DMF belirleme yöntemleri ile grup farklılıkları tespit edilir. DMF tek biçimli ve tek biçimli olmayan olmak üzere iki farklı şekilde görülebilir (Camilli ve Shepard, 1994; Mellenbergh, 1989). Tek biçimli DMF, madde yanıt fonksiyonlarının iki grup arasında aynı olmadığı ve herhangi bir çakışma olmadığı durumda gözlenir, yani tüm yetenek düzeylerinde (θ) bir grubun daha yüksek tepki vermesinin daha olası olması durumudur. Eğer madde yanıt fonksiyonları iki grup için çakışırsa tek biçimli olmayan DMF görünür (Camilli ve Shepard, 1994). Bir teste DMF gösteren maddelerin bulunması yanlılığa neden olacağından testin grup karşılaştırmalarına ilişkin geçerliğini düşürebilir (Dorans ve Holland, 1993). Yanlılık bir sistematik hata nedeni olduğundan, sistematik hatanın artması yalnızca gruplar arası farkların belirlenebilirliğini azaltmaz, aynı zamanda farkların yönü ve büyüklüğünün yanlış yorumlanmasına da neden olabilir (OECD, 2015). Dolayısıyla, PISA gibi uluslararası değerlendirme araştırmalarının temelde ülkeler arası yeterlikleri karşılaştırdığı düşünüldüğünde, dilsel ve kültürel olarak DMF gösteren maddeleri içermemesinin sağlanması gerekmektedir.

Alan yazında pek çok çalışmada farklı kültürlerle ve dillerle yapılan değerlendirmelerde DMF'nin ortaya çıktığı bulunmuştur (Allalouf, Hambelton, ve Sireci, 1999; Budgell, Raju, ve Quartetti, 1995; Ercikan ve Koh, 2005; Grisaya ve Monseur, 2007; Huang, Wilwon ve Wang, 2014; Sireci ve Swaminathan, 1996; Wu ve Ercikan, 2006). Örneğin, Grisaya ve Monseur (2007) çalışmalarında PISA 2000 uygulaması kapsamında yer alan okuma maddelerinin DMF gösterip göstermediğini 47 ülke arasında incelemişlerdir. Araştırmanın bulguları, çevirinin DMF'nin ortaya çıkmasında çok etkili olduğunu ve farklı diller arasında ortaya çıkan DMF'li madde sayısının aynı dili konuşan farklı kültürler arasındaki DMF'li madde sayısından çok daha fazla olduğunu göstermiştir. Benzer şekilde Asil ve Gelbal (2012), PISA 2006 kapsamında yer alan öğrenci anketlerinin kültürler ve diller arası eşdeğerliğini Avustralya, Yeni Zelanda, Amerika Birleşik Devletleri ve Türkiye örneklemi üzerinde incelemişlerdir. Araştırmanın bulguları ülkeler arasında dil değişikçe ve kültürel özellikler farklılaştıkça DMF'li madde sayılarının değiştiğini göstermiştir. Huang, Wilson ve Wang (2014) PISA

2006 fen uygulamasının kültürler ve diller arasında DMF gösterip göstermediğini Birleşik Devletler ve Kanada, Çin-Hong Kong ve Çin, Birleşik Devletler ve Çin ülkeleri arasında incelemişlerdir. Araştırmanın sonuçları en çok DMF gösteren maddenin Birleşik Devletler ve Çin arasında ortaya çıktığını, İngilizce konuşan Birleşik Devletler ve Kanada arasında ise DMF'nin göz ardı edilebilir düzeyde olduğunu göstermiştir. Bu çalışmalar dil farklılığının DMF'nin ortaya çıkmasında oldukça etkili bir faktör olduğunu göstermektedir.

Bu çalışmalara ek olarak alan yazında kültürler ve diller arasında ortaya çıkan DMF'yi etkileyebileceği düşünülen öğrenci özelliklerinden kaynaklanabilecek etmenler de ele alınmıştır. Wu ve Ercikan (2006) TIMSS 1999 uygulaması kapsamında Birleşik Devletler ve Tayvan arasında maddelerin DMF gösterip göstermediğini lojistik regresyon yöntemi ile incelemişlerdir. Çalışmalarında öğrencilerin okul dışı fazladan çalışma saatlerinin bir DMF kaynağı olup olmamasının yanı sıra, bu durumun ülkeler arasında ortaya çıkan DMF'li madde sayılarına bir etkisinin olup olmadığını incelemişlerdir. Araştırmanın bulguları okul dışı fazladan çalışma saatinin bir DMF kaynağı olduğunu ve modele eklendiğinde ülkeler arasında ortaya çıkan DMF'li madde sayısını %29 oranında azaldığını göstermiştir. Husin (2014) PISA 2012 matematik maddelerinin Malezya örneğinde ana dilde ve İngilizce uygulamaları arasında DMF gösterip göstermediğini, ayrıca DMF'nin varlığının öğrencilerin okuma becerilerinden ve sosyoekonomik düzeylerinden (SED) etkilenip etkilemediğini lojistik regresyon yöntemi ile incelemiştir. Araştırmanın bulguları testin uygulandığı dilin yanı sıra madde özelliklerinin, SED'in ve okuma becerisinin de genel test performansını etkilediğini ve modele bu değişkenlerin eklenmesinin DMF gösteren madde sayısında azalmaya neden olduğunu göstermiştir.

DMF'nin önemli düzeyde ortaya çıktığı bir diğer grup ise cinsiyet gruplarıdır. Alan yazında pek çok çalışmada (Atalay Kabasakal ve Kelecioğlu, 2012; Lan, 2014; Le, 2009; Lyons-Thomas, Sandilands ve Ercikan, 2014) uluslararası uygulanan sınavlarda cinsiyet grupları arasında DMF ortaya çıktığı raporlanmıştır. Bunlara ek olarak cinsiyete ilişkin DMF analizlerinde madde özelliklerinin DMF'nin yönüne etkisine ilişkin araştırmalar da yer almaktadır. Lyons-Thomas, Sandilands ve Ercikan (2014) çalışmalarında PISA 2009 uygulamasında yer alan matematik maddelerinin Kanada, Çin-Şangay, Finlandiya ve Türkiye örneklerinde cinsiyete göre DMF gösterip göstermediğini madde türleri bazında incelemiştir. Araştırmanın bulgularına göre genel olarak çoktan seçmeli maddeler erkekler lehine, açık uçlu maddeler ise kadınlar lehine DMF göstermiştir. Benzer şekilde Le (2009) PISA 2006 deneme uygulamasında yer alan fen maddelerinin cinsiyetler arasında DMF gösterip göstermediğini ülkelere ve madde özelliklerine göre incelemiştir. Araştırmada madde özellikleri olarak PISA çerçevesinde belirlenen odak, içerik, yeterlik ve bilimsel bilgi olmak üzere dört madde özelliği ve madde türleri temel alınmıştır. Araştırmanın bulgularına göre çoktan seçmeli ve kapalı uçlu maddeler, evrensel odaklı maddeler, çevre ile ilgili maddeler ve bilimsel bir olguyu açıklamayı gerektiren maddeler çoğunlukla erkekler lehine DMF göstermektedir. Diğer yandan doğal kaynaklarla ilgili maddeler, bilimsel bilgiyi tanımlama, bilimsel kanıt ve veri kullanma gerektiren maddeler ise kadınlar lehine DMF göstermiştir.

Alan yazın genel olarak incelendiğinde uluslararası uygulamalarda kültüre, dile ve cinsiyete dayalı DMF çalışmalarının çok sayıda olduğu görülmektedir. Özellikle PISA uygulamalarının uluslararası geçerliğini test etmeye yönelik ülkeler arasında DMF ve ölçme değişmezliği çalışmaları oldukça yaygındır. Öte yandan PISA 2015 teknik raporunda ülkeler arası DMF'nin nasıl belirlendiği ve DMF'li maddelerin nasıl kontrol edildiğine ilişkin bilgilere yer verilmiştir (OECD, 2015). Bu bağlamda OECD'nin PISA 2015 uygulamasında ülkelere göre olası yanlılıkları göz önünde bulundurduğu sonucuna varılabilir. Diğer yandan raporda cinsiyete yönelik herhangi bir DMF çalışmasından bahsedilmemektedir. Ayrıca alan yazında cinsiyete yönelik DMF çalışmalarında çoğunlukla madde türünün cinsiyet DMF'sinin yönü ile etkileşiminden bahsedilmiş; ancak kültüre yönelik DMF çalışmalarında olduğu gibi öğrenci özelliklerinin (SED, okuma becerisi gibi) cinsiyete göre DMF'ye etkisine ilişkin bir çalışmaya rastlanmamıştır. Bu nedenlerle bu araştırmada cinsiyete dayalı DMF ve DMF'yi etkileyebilecek olası öğrenci özellikleri farklı ülkeler bağlamında incelenmiştir. Araştırma kapsamında cinsiyete dayalı DMF'ye SED ve okuma becerisinin etkisi incelenmiştir.

PISA 2015'te SED; aile eğitimi, aile mesleği ve evdeki olanaklar değişkenlerinden temel bileşenler analiziyle oluşturulan bir indekstir (OECD, 2015). SED'in öğrencilerin eğitimsel kazanımları elde etmesinde ve akademik başarıları üzerinde önemli bir belirleyici faktör olduğu düşünülmektedir (Barr, 2015; Coleman ve arkadaşları, 1966; Hecht, Burgess, Torgesen, Wagner ve Rashotte, 2000; White, 1982). Yüksek SED ve düşük SED'in karşılaştırıldığı bazı çalışmalarda eğitimsel çıktıların zayıflığının nedenleri arasında aile eğitim düzeyi düşüklüğü, okul kaynaklarının azlığı, aile katılımının azlığı bulunmuştur (Schmidt, Cogan, & McKnight, 2011). PISA uygulamalarında SED bakımından avantajlı olan öğrencilerin ve okulların dezavantajlı akranlarına göre çok daha yüksek puanlar aldığı gözlenmiştir (OECD, 2015). Her ne kadar düşük SED doğrudan düşük başarıya neden olur gibi bir çıkarım yapmak doğru olmasa da, sosyoekonomik düzeyin başarı üzerindeki etkisi yadsınamaz. Bu bağlamda cinsiyete bağlı DMF analizlerinde sosyoekonomik düzey değişkeninin modele dahil edilmesiyle sosyoekonomik düzeyinde bir DMF kaynağı olup olmadığının ve cinsiyete dayalı DMF'nin varlığını etkileyip etkilemediğinin değerlendirilmesine olanak sağlamaktadır.

PISA, ölçtüğü temel alanın yanı sıra öğrencilerin günlük hayatta karşılarına çıkabilecek çeşitli yazılı materyalleri anlama ve yorumlama düzeylerini de değerlendirir. PISA'da okuma becerilerinde ele alınan bilişsel yeterlikler; bir metni basit olarak çözümlemenin ötesinde metni çözümlerken uygun stratejileri kullanma becerisi ve bu becerinin farkında olma gibi üst düzey bilişsel becerileri de içine alır. Günümüz bilgi çağında bireyin okuma, okuduğunu anlama, yorumlama, muhakeme etme gibi yeterliklere sahip olması beklenmektedir. Öğrencilerin okuma becerilerinin fen ve matematik başarılarını etkilediği bilinmektedir (Demps ve Onwuegbuzie, 2001; Nolen, 2003; O'Reilly & McNamara, 2007). Cromley (2009) 32 ülke arasında okuma ve fen başarısı arasında ortalama korelasyon katsayılarını PISA 2000 verisinde 0,84, PISA 2003 verisinde 0,805 ve PISA 2006 veri setinde 0,819 bulmuştur. Her üç veri setinde de okuma-fen arasındaki korelasyon katsayıları ülkelere göre değişmekle birlikte, düşük okuma puanı olan ülkelerde en düşük katsayılar elde edilmiştir. Çoğu öğretmen ve program geliştirme uzmanı bazı matematik maddelerinin öğrencilerin doğru olarak cevaplayabilmeleri için yüksek düzeyde okuma becerisine sahip olması gerektiği belirtmektedir (NCES, 2003). Büyük ölçekli sınavlarda genellikle fen ve matematik maddelerinin cevaplanabilmesi için ciddi bir okuma yükü gerekir. Eğer öğrencinin okuma becerisi düşük bir düzeydeyse öğrenci kendi gerçek başarısını gösteremeyebilir. Diğer bir ifadeyle öğrenciler verilen problem durumunu doğru bir şekilde matematiksel formata dönüştürmede başarısız olursa kendi performansını gösteremeyecektir. Walker, Zhang ve Surber (2008) yaptıkları çalışmada çok düzeyli DMF çerçevesinde öğrencilerin matematik maddelerindeki performansının okuma başarıları ile ilişkili olduğunu, düşük okuma becerisine sahip öğrencilerin matematik maddelerini doğru cevaplama olasılıklarının düşük olduğunu bulmuşlardır. Alanyazına dayanarak cinsiyete bağlı DMF analizlerinde okuma başarısı değişkeninin modele dahil edilmesiyle bu değişkenin bir DMF kaynağı olup olmadığı ve cinsiyete dayalı DMF'nin varlığını etkileyip etkilemediği incelenmiştir.

PISA 2015 uygulamasında fen bilimleri temel alan olarak alınmıştır. Bu bağlamda uygulamada yer alan okuma ve matematik testlerinde yer alan maddeleri önceki yıllarda uygulanan maddelerden oluşurken, fen testlerinde hem önceki yıllarda kullanılan maddeler hem de ilk defa uygulanan yeni maddeler yer almaktadır. PISA 2015 uygulamasında diğer yıllardan farklı olarak madde kümeleri yer almaktadır ve bu kümeler 96 farklı forma dağıtılmıştır. Ayrıca ülkelerin bir kısmına kağıt kalem formları uygulanırken diğer kısmında uygulama bilgisayar tabanlı yapılmıştır. Formların ilk 30'u kağıt kalem testi şeklinde geri kalanı bilgisayar tabanlı uygulamalarda kullanılmıştır. Uygulamada 18 fen bilimleri madde kümesi (S1-S18) yer almaktadır ve bu kümelerin 6 tanesi (S1-S6) daha önce uygulanmış maddelerden, 12 tanesi ise (S7-S18) yeni uygulanan maddelerden oluşmaktadır (OECD, 2015). Araştırma kapsamında bilgisayar tabanlı uygulamada yer alan ve ilk defa uygulanan maddelerden oluşan S12 fen bilimleri madde kümesi ele alınmıştır. Söz konusu maddelerin cinsiyete bağlı DMF gösterip göstermediği dokuz ülke kapsamında incelenmiştir. Ülkeler seçiminde 2015 uygulaması fen bilimleri başarı sıraları bir ölçüt olarak kullanılmış ve OECD ortalaması, ortalama üzeri ve ortalama altı olmak üzere her bir düzeyden üçer ülke araştırmaya dahil edilmiştir. Bu bağlamda araştırma OECD ortalaması üzerinde yer alan Japonya, Finlandiya ve Hong Kong; ortalamada yer alan Birleşik Devletler, Fransa ve İsveç; ortalama altında ise Türkiye, Brezilya ve İsrail ülkeleri ile gerçekleştirilmiştir.

Araştırmanın Amacı

Araştırmanın amacı PISA 2015 uygulaması fen bilimleri testinde yer alan S12 madde kümesinin farklı kültürler içerisinde cinsiyete bağlı DMF gösterip göstermediğinin incelenmesi; maddelerin bilgi ve içerik alanları ile cinsiyet DMF'si arasındaki ilişkinin incelenmesi ve bununla beraber öğrencilerin sosyoekonomik düzeylerinin ve okuma becerilerinin cinsiyete dayalı DMF'yi etkileyebilecek birer DMF kaynağı olup olmadığının belirlenmesidir. Bu kapsamda araştırmada aşağıdaki araştırma sorularına cevap aranmaktadır.

1. S12 madde kümesinde yer alan fen bilimleri maddeleri Japonya, Finlandiya, Hong Kong, Birleşik Devletler, Fransa, İsveç, Türkiye, Brezilya ve İsrail verilerinde cinsiyete göre DMF göstermekte midir?
2. Maddelerin bilgi ve içerik alanları ile cinsiyet DMF'si arasında bir ilişki var mıdır?
3. Her bir ülke içinde öğrencilerin sosyoekonomik düzeylerinin grup üyeliğinin ötesinde DMF'ye bir etkisi var mıdır?
4. Her bir ülke içinde öğrencilerin okuma becerilerinin grup üyeliği ve sosyoekonomik düzeyin ötesinde DMF'ye bir etkisi var mıdır?

DMF belirleme yöntemleri

DMF belirlemede çok sayıda yöntem geliştirilmiştir. Bu yöntemler arasında Mantel-Haenszel yöntemi (Holland ve Thayer, 1988), standartlaştırma yöntemi (Dorans ve Kulick, 1986), lojistik regresyon yöntemi (Swaminathan ve Rogers, 1990), SIBTEST yöntemi (Shealy ve Stout, 1993), Lord'un k-kare testi (Lord, 1980; Wright ve Stone, 1979), olabilirlik oranı testi (MTK-OO; Thissen, Steinberg, ve Wainer, 1988; Wang ve Yeh, 2003), çoklu göstergeler çoklu nedenler modeli (MIMIC; Finch, 2005; Oort, 1998) sayılabilir. MH, standartlaştırma yöntemi ve lojistik regresyon yöntemleri gözlenen puana dayalı yöntemlerdir. Yaygın olarak kullanılmalarına karşın örtük özellik yerine toplam puan kullanılmasının doğruluğu tartışmalıdır. Araştırmalar cevaplar iki ya da üç parametrelili MTK gibi daha karmaşık bir modelle üretildiğinde bu yöntemlerin DMF'yi yanlış belirleyebileceğini göstermiştir (Meredith & Millsap, 1992; Millsap & Meredith, 1992).

MIMIC yöntemi diğer yöntemlere göre daha yeni bir yöntemdir. MIMIC, DMF belirlemede kullanılan örtük özelliğe dayalı bir yapısal eşitlik modeli türüdür. Son yıllarda farklı koşullarda MIMIC yönteminin etkililiği ile ilgili çok sayıda çalışma yapılmış ve bu yöntem DMF belirleme çalışmalarında sıklıkla kullanılmaya başlanmıştır (Finch, 2005; Fleishman, Spector, & Altman, 2002; Gallo, Anthony, & Muthe'n, 1994; Glöckner-Rist & Hoijtink, 2003; Levine et al., 2003; MacIntosh & Hashim, 2003; Muthe'n, 1985; Muthe'n, Kao, & Burstein, 1991; Shih ve Wang, 2009; Oort, 1998).

MIMIC modelinin avantajlarından biri nedensel gösterge olarak tanımlanan en az bir gözlenen değişkenin örtük değişkeni yordamasıdır (Joreskog ve Goldberger, 1975). Ayrıca MIMIC dışsal değişkenlerle MTK modellerine genişletilebilen bir doğrulayıcı faktör analizi yöntemi olduğundan madde parametreleri kestirimi, DMF belirleme, MTK'nın tek boyutluluk ve koşullu bağımsızlığı gibi varsayımlarını esnetme gibi önemli madde analizi konularına da eş zamanlı olarak değinebilmektedir. Bunlara ek olarak MIMIC yönteminde gruplama değişkeni sürekli ya da süreksiz bir değer olabilir. Dolayısıyla, sadece tek bir süreksiz gruplama değişkeniyle DMF analizi yapmaya izin veren geleneksel DMF belirleme yöntemlerine (MH, SIBTEST ya da MTK-OO gibi) göre daha esnek olduğu söylenebilir (Wang, Shih ve Yang, 2009). MIMIC yönteminde çok boyutlu maddeler veya faktörler kolayca modellenebilir ve birden fazla grup arasında DMF incelenirken yorumlaması daha kolaydır. Ayrıca MIMIC modeli DMF test edilirken fazladan kovaryant değişkenlerin modele eklenmesine ve kovaryant değişkenlerin sürekli olmasına da olanak sağlar. Diğer yandan MIMIC yöntemi tek biçimli DMF'yi belirlemeye hassastır (Woods, 2009). Ancak son yıllarda tek biçimli olmayan DMF belirlemek için de MIMIC yönteminin kullanıldığı çalışmalar yapılmaktadır (Chun, 2014; Woods ve Grimm, 2011).

Bu çalışmada DMF analizleri MIMIC yöntemi ile gerçekleştirildiğinden yalnızca bu yöntemin hesaplanmasına detaylı olarak yer verilmiştir.

MIMIC ile DMF Analizi

MIMIC modeli hem ölçme modelinden hem de yapısal modelden oluşur. MIMIC yönteminin ölçme modeline ilişkin denklem aşağıdaki gibidir:

$$y_i^* = \lambda_i \eta + \beta_i z_k + \varepsilon_i, \quad (1)$$

Denklem (1)'de y_i^* , i maddesi için testin ölçmeyi amaçladığı θ için gizil cevap değişkeni, z cinsiyet, etnik köken gibi faktör analitik modelde yer alan DMF ile ilişkili olduğu düşünülen grupta değişkeni, λ_i ise faktör yüküdür ve MTK kapsamında i maddesi için eğitim parametresidir. ε_i , tesadüfi hatadır ve normal dağılıma sahiptir. β_i ise grup değişkeni ve z 'nin y_i^* üzerindeki etkisidir. $\beta_i = 0$ ise i maddesi z grupta değişkenleri üzerinde homojendir yani i maddesinde DMF yoktur, eğer $\beta_i \neq 0$ ise z 'nin y_i^* üzerinde doğrudan etkisi vardır. Yani i maddesi z grupta değişkenine göre DMF göstermektedir. y_i doğrudan gözlenemediği için test maddeleri y_i^* 'yi ölçmek amacıyla y_i^* 'yi iki kategorili madde cevaplarına (y_i) dönüştürür:

$$y_i = \begin{cases} 1, & y_i^* > \tau_i \\ 0, & \text{diğer} \end{cases} \quad (2)$$

Denklem (2)'de τ_i , madde güçlüğüne ilişkin sınır parametresini göstermektedir.

Yapısal modelde, θ ile z grupta değişkeni $\theta = \gamma'z + \xi$ denkleminde olduğu gibi doğrudan ilişkilidir. γ' , θ üzerindeki grup farklılıklarını tanımlayan regresyon katsayılarının bir vektörüdür ve genellikle DMF analizlerinde etki olarak kullanılır (Ackerman, 1992; Camilli, 1993); ξ ortalaması sıfır olan z 'den bağımsız olarak normal dağıldığı varsayılan artıktır.

MTK'da üç parametrelilik için (Hambelton, Swaminathan ve Rogers, 1991), n bireyinin i maddesini doğru yanıtlama olasılığı aşağıdaki gibi tanımlanmıştır:

$$P(U_{ni} = 1 | \theta_n) = c_i + (1 - c_i) \frac{\exp(a_i(\theta_n - b_i))}{1 + \exp(a_i(\theta_n - b_i))} \quad (3)$$

Denklem (3)'de U_{ni} , n bireyinin i maddesine verdiği cevap değişkeni, θ_n , n bireyinin örtük özelliğinin düzeyi a_i , b_i ve c_i ise sırayla i maddesinin madde ayırıcılık, madde güçlük ve şans parametrelerini göstermektedir. Her bir i maddesi için $c_i = 0$ olduğunda üç parametrelilik modeli iki parametrelilik modeline dönüşmektedir. Her bir i maddesi için $c_i = 0$ ve a_i parametreleri maddeler arasında eşit olduğunda üç parametrelilik modeli bir parametrelilik modeline dönüşmektedir.

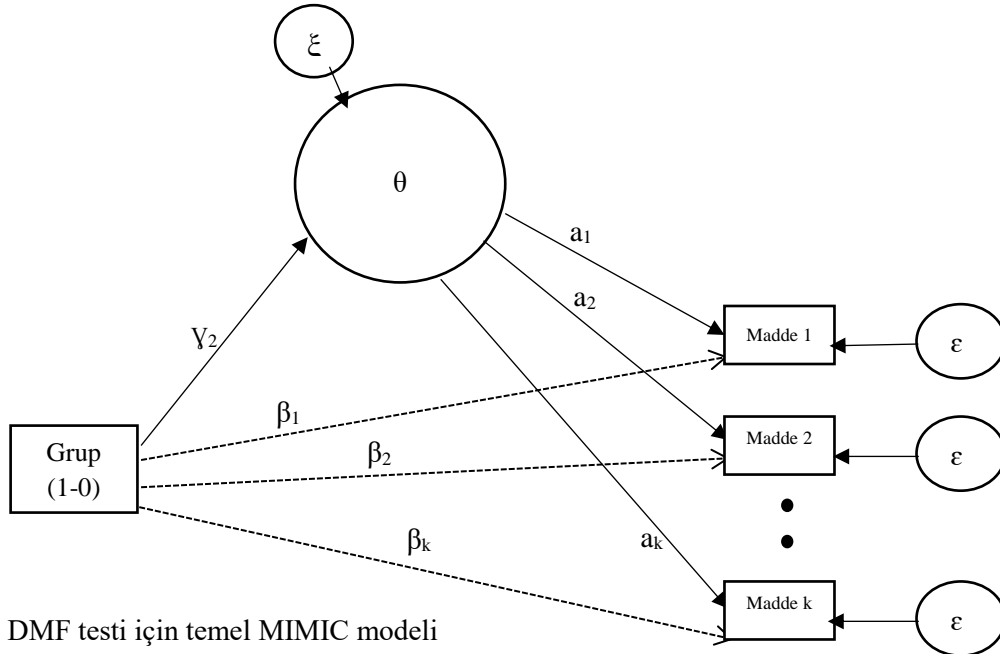
$$a_i^* = c(1 - \lambda_i^2 \psi)^{-1/2} \sigma_{00}^{1/2} \quad (4)$$

$$b_i^* = [(\tau_i - \beta_i' z) \lambda_i^{-1} - \mu_0] \sigma_{00}^{1/2} \quad (5)$$

Ψ , ξ 'nin varyansı; μ_0 ve σ_{00} ise θ 'nin ortalaması ve varyansıdır.

Referans grup için $z = 0$, odak grup için $z = 1$ olmak üzere tek bir grupta değişkeni olduğunda Denklem (5)'de görüleceği üzere gruplar arası madde güçlüğü farkı $\beta_i' \lambda_i^{-1} \sigma_{00}^{-1}$ 'e eşit olacaktır. σ_{00} , θ 'nin standart sapmasıdır ve birim olarak sınırlandırıldığında, $\beta_i' \lambda_i^{-1}$ tek biçimli DMF'nin etki büyüklüğüdür (Shih ve Wang, 2009). DMF testi için temel MIMIC modeli Şekil 1'de gösterilmiştir.

Şekil 1, madde cevaplarının DMF testi için grupta değişkeni üzerinde regresyonunu gösteren standart bir tek boyutlu MTK modelidir. Bu modelde θ üzerindeki farklılıklar kontrol altına alındığında eğer grupta değişkeninin madde cevaplarını yordaması anlamlı ise DMF gözlenir. Bu model tek biçimli DMF modeli olduğu için madde ayırt edicilikleri dolaylı olarak değişmezdir (Woods, 2009).



Şekil 1. DMF testi için temel MIMIC modeli

γ : örtük değişkende ortalamalar arası farkı gösteren regresyon katsayısı, β_i : i maddesinin eşik parametresinin grup değişkeni arasında farkını gösteren regresyon katsayısı ($i=1,2..k$), a_i : ayırıcılık parametresi, ϵ_i : i maddesi için ölçme hatası, ξ : θ için artık. (Woods, 2009, s.5)

DMF belirleme çalışmalarında referans ve odak grubu aynı ölçüğe yerleştirmek amacıyla eşleştirme değişkeni kullanılmaktadır. Bu sayede maddelerin DMF gösterip göstermediği değerlendirilir (Wainer, Sireci ve Thissen, 1991). Eşleşme değişkeni içsel ve ya dışsal olarak iki türlü seçilebilir. (Welch ve Miller, 1995). Eğer DMF analizi yapılacak maddeler aynı zamanda eşleşme değişkeni olarak kullanılıyorsa içsel değişken, eğer eşleşme değişkeni test dışından maddelerden oluşan bir set ise dışsal değişkendir. Pratikte, DMF analizi yapılan test ile aynı örtük özelliği ölçen DMF içermeyen madde setleri bulmak çoğunlukla zor olduğu için eşleşme değişkeni genellikle içsel olarak belirlenir (Shih ve Wang, 2009). Gerçek koşullarda testler kusursuz değildir ve DMF’li maddeler içerirler. Bu nedenle eşleştirme değişkenininin DMF içermeye ihtimali vardır. Ancak, DMF değerlendirmesinin doğru olarak yapılabilmesi için eşleşme değişkeninin DMF içermemesi gereklidir (Holland ve Thayer, 1988; Lord, 1980). Eğer DMF analizlerinde DMF içeren bir eşleşme değişkeni kullanılırsa, referans ve odak grupların performansı yanlı bir ölçüte dayalı olarak karşılaştırılacaktır ve sonuçlar yanıltıcı olacaktır. Diğer yandan eğer eşleşme değişkenleri DMF içermeyse, bu durumda da DMF analizi gerekli olmayacaktır. Dolayısıyla bu durum bir kısır döngü oluşturmaktadır (Shih ve Wang, 2009).

Finch (2005) MIMIC yöntemini iki parametrelili ve üç parametrelili MTK modellerinde tek biçimli DMF’yi belirlemede Mantel-Haenszel, MTK-OO ve SIBTEST yöntemleri ile karşılaştırmıştır. Yapılan çalışmada test uzunluğu, örneklem büyüklüğü, gruplar arası ortalama yetenek farkı, DMF etki büyüklüğü ve eşleşme değişkenindeki DMF düzeyi değişkenleri manipüle edilmiştir. Araştırma sonuçları MIMIC yönteminin diğer yöntemler kadar uygulanabilir olduğunu göstermiştir. Ayrıca eşleşme değişkenindeki DMF düzeyinden diğer yöntemler büyük ölçüde etkilenirken, MIMIC yönteminde sadece I.Tip hatada az miktarda azalma, güçte ise küçük bir düşme gözlenmiştir. Eşleşme değişkeninde DMF gösteren madde olması ihtimali olduğunda, MIMIC yöntemin DMF içeren maddelerden diğer yöntemlere göre daha az etkilenmesi nedeniyle kullanılmasının uygun olduğunu belirtmiştir.

Bir testte bir ya da daha fazla sayıda DMF içeren maddenin varlığı durumunda testteki diğer maddeler için DMF araştırmalarının sonuçları doğru olmayabilir. Örneğin; DMF içermeyen maddeler yanlışlıkla

DMF'li olarak saptanabilir. Bu durum testin 1. tip hatasını istenmeyen şekilde artırır (Clauser, Mazor ve Hambleton, 1993). İçsel eşleşme değişkenindeki DMF etkisini sınırlamanın bir yolu ölçek arıtma yöntemi kullanmaktır. Ölçek arıtma yöntemleri DMF belirleme yöntemlerine uyarlanmıştır. Örnek olarak iki-aşamalı ya da iteratif Mantel-Haenszel (Holland ve Thayer, 1988), iteratif Mantel yöntemi, iteratif genelleştirilmiş Mantel yöntemi (Wang ve Su, 2004a, 2004b), iteratif lojistik regresyon (French ve Maller, 2007), ve iteratif MTK yöntemi (Candell ve Drasgow, 1988) verilebilir. Aynı prensipteki ölçek arındırma yöntemi MIMIC için de aşağıdaki şekilde uyarlanmıştır:

1. MIMIC ile 1. madde DIF açısından değerlendirilir ve testteki diğer tüm maddeler (DMF analizlerinin yürütüldüğü maddeler) çalışan madde olarak kullanılır. İlk adım testteki tüm maddeler için tekrarlanır.
2. Önceki adımlarda DMF içermeyen maddeler çalışan madde olarak tanımlanır ve tüm maddeler için DMF analizi yapılır.
3. İki ardışık iterasyonda aynı DMF'li madde kümesi belirleninceye kadar 3. adım tekrar edilir (Shih ve Wang, 2009).

YÖNTEM

Araştırma kapsamında PISA 2015 uygulaması S12 madde kümesinde yer alan fen maddelerinin cinsiyet grupları arasında DMF gösterip göstermediği ve cinsiyete dayalı DMF'yi SED ve okuma başarısı değişkenlerinin etkileyip etkilemediği incelenmiştir. Bu bağlamda araştırma PISA 2015 uygulaması fen maddelerinin geçerliğini test etmeye yönelik betimsel bir araştırmadır.

PISA 2015 uygulamasında yer alan fen maddeleri yeterlik, bilgi ve içerik olmak üzere üç boyutta sınıflandırılmıştır. Yeterlilik bağlamında maddeler bilimsel araştırma desenleme ve değerlendirme (Evaluate and design scientific enquiry), olguları bilimsel yollarla açıklama, veri ve kanıtları bilimsel olarak yorumlama olmak üzere üç grupta sınıflandırılmıştır. Benzer şekilde bilgi bağlamında alan (content), epistemik ve süreç bilgisi olmak üzere üç grupta; içerik bağlamında dünya ve uzay, yaşam ve fiziksel olmak üzere üç grupta sınıflandırılmıştır (OECD, 2015). Araştırma kapsamına dahil edilen S12 madde kümesinde beş başlıkta 17 madde yer almaktadır. Maddelerin yeterlilik, bilgi ve içerik alanına göre dağılımları Tablo 1'de verilmiştir.

Tablo 1. S12 Fen Maddelerinin Özellikleri

Konu	Madde kodları*	Yeterlik	Bilgi	İçerik
Göktaşları ve Kraterler	CS641Q01S-M1	Olguları bilimsel yollarla açıklama	Alan	Fiziksel
	CS641Q02S-M2	Olguları bilimsel yollarla açıklama	Alan	Dünya ve Uzay
	CS641Q03S-M3	Veri ve kanıtları bilimsel olarak yorumlama	Alan	Dünya ve Uzay
	CS641Q04S-M4	Veri ve kanıtları bilimsel olarak yorumlama	Alan	Dünya ve Uzay
Deniz ortamında sesler	CS626Q01S-M5	Olguları bilimsel yollarla açıklama	Alan	Fiziksel
	CS626Q02S-M6	Bilimsel araştırma desenleme ve değerlendirme	Süreç	Fiziksel
	CS626Q03S-M7	Veri ve kanıtları bilimsel olarak yorumlama	Süreç	Fiziksel
	DS626Q04C-M8	Veri ve kanıtları bilimsel olarak yorumlama	Süreç	Yaşam
Eğim-Yüzey İncelemesi	DS637Q01C-M9	Olguları bilimsel yollarla açıklama	Epistemik	Dünya ve Uzay
	CS637Q02S-M10	Olguları bilimsel yollarla açıklama	Epistemik	Dünya ve Uzay
	DS637Q05C-M11	Veri ve kanıtları bilimsel olarak yorumlama	Epistemik	Dünya ve Uzay
Beyin Kontrollü Robotik	DS610Q01C-M12	Olguları bilimsel yollarla açıklama	Alan	Yaşam
	CS610Q02S-M13	Olguları bilimsel yollarla açıklama	Alan	Yaşam
	CS610Q04S-M14	Bilimsel araştırma desenleme ve değerlendirme	Alan	Yaşam
Sürdürülebilir Balık Çiftliği	CS601Q01S-M15	Olguları bilimsel yollarla açıklama	Alan	Yaşam
	CS601Q02S-M16	Veri ve kanıtları bilimsel olarak yorumlama	Alan	Yaşam
	CS601Q04S-M17	Olguları bilimsel yollarla açıklama	Alan	Fiziksel

*Madde kodları araştırmacılar tarafından M1-M17 şeklinde tekrar kodlanmıştır.

Örneklem

Araştırma kapsamında seçilen fen bilimleri uygulaması S12 madde kümesinde yer alan maddelerin cinsiyete göre DMF gösterip göstermediği Japonya, Finlandiya, Hong Kong, Birleşik Devletler, Fransa, İsveç, Türkiye, Brezilya ve İsrail örneklemelerinde incelenmiştir. Dolayısıyla araştırmanın evrenini söz konusu dokuz ülkedeki 15 yaş grubu öğrencileri oluşturmaktadır.

Araştırmanın örneklemini dokuz ülkede PISA uygulamasına katılan ve S12 madde kümesinin dahil olduğu kitapçıkları cevaplayan araştırmanın kapsamı dahilinde ele alınan değişkenlerden eksik verisi bulunmayan öğrenciler oluşturmaktadır. Öğrencilerin ülkelere göre dağılımları Tablo 2’de özetlenmiştir. PISA uygulamalarında örnekleme yöntemi olarak tabakalı örnekleme kullanılmaktadır ve uygulama formları öğrencilere rastgele olarak dağıtılmaktadır. Tüm ülkeler için madde kümesinin uygulandığı örneklem rastgele olduğundan belirli bir madde kümesinin uygulandığı bireyleri örnekleme dahil etmek herhangi bir örnekleme yanlılığı oluşturmayacaktır.

Tablo 2. Ülkelere Ve Cinsiyetlere Göre Öğrenci Frekansları

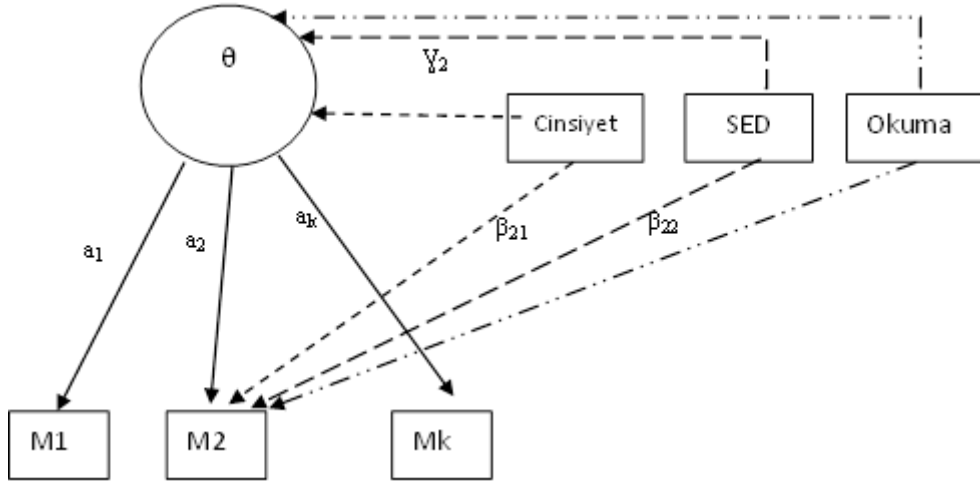
Ülke	Kadın		Erkek		Toplam	
	f	%	f	%	f	%
Japonya	600	48,7	632	51,3	1232	100
Finlandiya	599	48,4	639	51,6	1238	100
Hong Kong	495	50,3	490	49,7	985	100
Birleşik Devletler	505	50,2	500	49,8	1005	100
Fransa	557	50,5	546	49,5	1103	100
İsveç	469	50,3	464	49,7	993	100
Türkiye	511	49,9	513	50,1	1024	100
Brezilya	1281	51,0	1232	49,0	2513	100
İsrail	654	57,1	492	42,9	1146	100
Toplam	5671	0,51	5508	0,49	11239	100

Tablo 2 incelendiğine S12 madde kümesinin uygulandığı örneklem büyüklüğünün ülkeler arasında genel olarak benzer dağıldığı görülmektedir. Yalnızca Brezilya örneklemini diğer ülkelerden daha büyük görmektedir. Genel olarak incelendiğinde İsrail hariç ülkelerde kadın ve erkek grupları arasında fark en fazla %2 kadardır. Dolayısıyla örneklem büyüklükleri cinsiyet grupları arasında benzerdir.

Verilerin Analizi

Verilerin analizinde cinsiyet grupları arasında olası DMF içeren maddeleri belirlemeye yönelik MIMIC yöntemi kullanılmıştır. Analizler iki aşamada gerçekleştirilmiştir. İlk aşamada maddelerin her bir ülke içinde cinsiyet grupları arasında DMF gösterip göstermediği incelenmiştir. Analizlerde odak grup olarak kızlar alınmıştır. Daha sonra modele sırasıyla SED ve okuma becerisi değişkenleri eklenmiş ve söz konusu değişkenlerin cinsiyete bağlı DMF’yi nasıl etkilediği incelenmiştir. Araştırma kapsamında SED değişkeni olarak PISA kapsamında hesaplanan indeks değişkeni kullanılmıştır. Öğrencilerin okuma becerilerine yönelik ise uygulama kapsamında okuma testi sonucunda hesaplanan okuma becerisi puanları kullanılmıştır.

MIMIC yöntemi ile DMF analizleri, maddelerin örtük özelliği gruplar arasında eşit düzeyde ölçüp ölçmediğini belirlemek amacıyla tam model ile sınırlandırılmış model arasındaki uyumun karşılaştırılmasını içermektedir. Model karşılaştırmalarında sınırlandırılmış temel model kullanılmıştır. Araştırma kapsamında ele alınan tam model Şekil 2’de verilmiştir.



Şekil 2. Sınırlandırılmış temel model yaklaşımına yönelik tek biçimli DMF için tam model

Verilerin analizinde öncelikle maddelerin cinsiyet gruplarına göre DMF gösterip göstermediği incelenmiştir. Sonrasında DMF gösteren maddeler için modele öncelikle SED, ardından okuma becerisi değişkenleri eklenmiştir. DMF’li maddeler belirlenirken ilk olarak modelde cinsiyet değişkeninin manidarlığına bakılmıştır. Sonrasında ise tam model ile sınırlandırılmış model uyumları arasında anlamlı bir farklılık bulunup bulunmadığı incelenmiştir. Veriler 1-0 şeklinde puanlandığı için parametre kestirimlerinde ağırlıklandırılmış en küçük kareler (WLS) yöntemi kullanılmıştır. WLS kestirimlerinde model karşılaştırmalarında ki-kare değerleri arasındaki farklar, ki-kare dağılımı göstermediğinden ki-kare fark testi yapılmaz. Bu nedenle model uyumları arasındaki fark değerlendirilirken program tarafından sağlanan DIFFTEST fark testi kullanılmıştır (Muthen ve Muthen, 2010).

BULGULAR

Araştırmanın birinci alt problemine yönelik PISA 2015 fen bilimleri uygulamasında yer alan S12 madde kümesinde yer alan maddelerin Japonya, Finlandiya, Hong Kong, Birleşik Devletler, Fransa, İsveç, Türkiye, Brezilya ve İsrail örneklerinde cinsiyete göre DMF gösterip göstermediği incelenmiştir. DMF analizi bulguları Tablo 3’te özetlenmiştir.

Tablo 3. S12 Madde Kümesinde Yer Alan Soruların Cinsiyete Göre DMF Analizi Sonuçları

		Göktaşları ve Kraterler				Deniz Ortamında Sesler				Eğim-Yüzey İncelemesi			Beyin Kontrollü Robotik			Sürdürülebilir Balık Çiftliği		
		M01	M02	M03	M04	M05	M06	M07	M08	M09	M10	M11	M12	M13	M14	M15	M16	M17
Japonya	β	0,149	0,5	-0,346	-0,309	0,226	0,158	-0,09	-0,381	-0,06	0,074	-0,033	0,201	0,036	-0,088	0,104	-0,188	0,095
	p	0,154	0,000	0,069	0,232	0,004	0,071	0,4	0,000	0,445	0,327	0,786	0,517	0,781	0,356	0,478	0,126	0,266
Hong Kong	β	0,031	0,474	-0,315	0,036	0,367	-0,012	-0,021	-0,243	-0,038	-0,011	0,051	0,084	0,075	-0,154	0,485	-0,165	-0,069
	p	0,684	0,000	0,013	0,851	0,000	0,919	0,83	0,035	0,678	0,892	0,618	0,356	0,696	0,171	0,008	0,365	0,501
Finlandiya	β	0,192	0,113	-0,13	-0,774	0,259	0,219	0,108	-0,473	-0,247	-0,068	0,251	0,142	-0,042	-0,006	0,276	-0,258	-0,108
	p	0,036	0,275	0,363	0,425	0,030	0,067	0,451	0,006	0,12	0,848	0,003	0,689	0,798	0,954	0,015	0,051	0,649
Birleşik Devletler	β	0,256	-0,577	-0,159	0,007	0,303	-0,138	0,148	-0,487	-0,205	0,089	0,132	-0,057	-0,157	0,187	0,492	-0,095	-0,259
	p	0,005	0,24	0,327	0,965	0,001	0,184	0,386	0,075	0,111	0,27	0,23	0,596	0,296	0,099	0,010	0,351	0,249
Fransa	β	0,204	0,068	-0,331	-0,109	0,135	0,062	-0,163	-0,064	-0,416	-0,233	0,442	0,226	0,146	-0,171	0,142	-0,156	0,198
	p	0,196	0,674	0,141	0,694	0,093	0,472	0,4	0,571	0,462	0,003	0,013	0,250	0,295	0,286	0,39	0,265	0,414
İsveç	β	0,168	0,102	0,035	-0,465	0,354	-0,055	0,064	-0,469	-0,246	-0,096	0,298	3,412	0,408	0,245	0,081	0,08	-0,323
	p	0,061	0,387	0,812	0,105	0,000	0,646	0,587	0,009	0,087	0,342	0,008	0,413	0,01	0,008	0,57	0,745	0,498
Türkiye	β	0,217	0,183	-0,195	-0,235	0,302	0,007	0,029	-0,35	0,04	0,019	0,015	0,05	-0,014	-0,004	0,276	-0,046	-0,212
	p	0,005	0,027	0,033	0,006	0,000	0,926	0,704	0,005	0,798	0,851	0,872	0,874	0,896	0,972	0,245	0,59	0,726
Brezilya	β	0,067	0,172	-0,049	-0,595	0,163	0,09	-0,024	-0,296	0,193	-0,016	-0,02	0,095	-0,021	-0,077	0,181	-0,01	-0,069
	p	0,153	0,000	0,674	0,322	0,002	0,119	0,703	0,003	0,461	0,766	0,75	0,232	0,833	0,373	0,176	0,899	0,369
İsrail	β	0,29	-0,031	-0,244	-0,168	0,149	-0,192	0,026	0,061	-0,023	-0,05	0,157	0,281	-0,36	0,033	0,444	-0,247	0,016
	p	0,002	0,803	0,118	0,28	0,063	0,125	0,816	0,547	0,827	0,554	0,223	0,004	0,035	0,793	0,008	0,012	0,87

*Konu başlıkları araştırmacılar tarafından çevrilmiştir.

Tablo 3 incelendiğinde Türkiye verisinde 6, Hong Kong, İsveç ve İsrail verilerinde 5'er, Finlandiya verisinde 4, Japonya, Birleşik Devletler ve Brezilya verilerinde 3'er, Fransa verisinde 2 maddede cinsiyete dayalı DMF ortaya çıktığı görülmektedir. Genel olarak incelendiğinde 17 maddeden 13'ü farklı ülkelerde cinsiyete göre DMF göstermiştir. Madde özellikleri incelendiğinde olguları bilimsel yollarla açıklama yeterliği gerektiren 1., 2., 5., 12., ve 15. maddeler ve bilimsel araştırma desenleme ve değerlendirme yeterliği gerektiren 14. madde erkekler lehine, veri ve kanıtları bilimsel olarak yorumlama yeterliği gerektiren 3., 4., 8. ve 16. maddeler kadınlar lehine DMF göstermiştir. Bu bulgular daha önce Le (2009) tarafından yapılan çalışma ile benzerlik göstermektedir. Diğer yandan olguları bilimsel yollarla açıklama yeterliği gerektiren 10. madde yalnızca Fransa örnekleminde kadınlar lehine DMF göstermiştir. 12. madde ise İsveç örnekleminde erkekler lehine DMF gösterirken İsrail örnekleminde kadınlar lehine DMF göstermiştir. Benzer şekilde veri ve kanıtları bilimsel olarak yorumlama yeterliği gerektiren 11. madde Fransa ve İsveç örneklemlerinde erkekler lehine DMF göstermiştir. Cinsiyete göre DMF'nin anlamlı olduğu her bir madde için modele SED değişkeni eklenmiş ve SED değişkeninin anlamlı etkisinin olduğu analiz sonuçları Tablo 4'te verilmiştir.

Tablo 4. Cinsiyet ve SED DMF Modeli Sonuçları

			Türkiye	İsveç	
			M04	M11	M13
Cinsiyet Modeli	Cinsiyet	β	-0,235	0,298	0,408
		p	0,006	0,008	0,010
Cinsiyet ve SED Modeli	Cinsiyet	β	-0,196	0,317	0,308
		p	-0,231	0,180	0,104
Cinsiyet ve SED Modeli	SED	β	0,006	-0,254	-0,213
		p	0,031	0,728	0,146

Tablo 4 incelendiğinde modele SED değişkeni eklendiğinde cinsiyete göre DMF gösteren maddelerden Türkiye ve İsveç örneklemlerinde yer alan 3 maddenin etkilendiği gözlenmektedir. 4. madde başlangıçta yalnızca Türkiye örnekleminde kadınlar lehine DMF göstermektedir. Modele SED değişkeninin eklenmesiyle söz konusu maddenin cinsiyete ilişkin DMF'sinin ortadan kalktığı ve SED değişkenine dayalı bir DMF'nin ortaya çıktığı görülmektedir. 11. ve 13. maddeler ise başlangıçta İsveç örnekleminde erkekler lehine DMF göstermektedir. Modele SED değişkeninin eklenmesiyle söz konusu maddelerde cinsiyete dayalı DMF ortadan kalkmıştır. Her bir ülke için modele SED değişkeninin yanı sıra okuma değişkeni eklendiğinde anlamlı etki elde edilen analiz sonuçları ise Tablo 5'te verilmiştir.

Tablo 5. Cinsiyet, SED ve Okuma Becerisi DMF Modeli Sonuçları

			Türkiye	Finlandiya	Hong Kong	
			M03	M08	M03	M08
Cinsiyet Modeli	Cinsiyet	β	-0.195	-0.473	-0.315	-0.243
		p	0.033	0.006	0.013	0.035
Cinsiyet, SED ve Okuma Modeli	Cinsiyet	β	-0.188	-0.817	-0.581	-0.5
		p	0.075	0.262	0.083	0.191
Cinsiyet, SED ve Okuma Modeli	SED	β	-0.088	-0.07	0.19	-0.231
		p	0.036	0.715	0.007	0.105
Cinsiyet, SED ve Okuma Modeli	Okuma	β	0.000	-0.008	-0.004	-0.006
		p	0.859	0.605	0.394	0.494

Tablo 5'te görüldüğü gibi modele SED değişkeni ile birlikte okuma becerisi değişkeni eklendiğinde başlangıçta Türkiye ve Finlandiya'da cinsiyet DMF'si gösteren birer maddenin, Hong Kong'da ise iki maddenin cinsiyete dayalı DMF göstermediği gözlenmektedir. 1. maddenin Türkiye ve Hong Kong örneklemelerinde SED değişkenine göre DMF gösterdiği görülmektedir. Genel olarak incelendiğinde maddelerin başlangıçta kadınlar lehine DMF gösterdiği modele okuma değişkeninin eklenmesiyle DMF'nin ortadan kalktığı görülmektedir. Alan yazında genel olarak kadınların okuma becerisi açısından erkeklerden daha yüksek puanlar aldığı raporlanmaktadır (Hyde ve Lin, 1988; Logan ve Johnstone, 2009). Bu bağlamda söz konusu maddelerde başlangıçta ortaya çıkan cinsiyet DMF'sinin okuma becerileri arasındaki farklardan kaynaklanabileceği sonucuna varılabilir. Söz konusu maddelerde yalnızca üç ülkede okuma becerisinin cinsiyet DMF'si üzerinde etkili olması ise ülkelerde konuşulan dillerin özelliklerinden kaynaklanabilir.

Bulgular genel olarak incelendiğinde söz konusu madde kümesi içinde maddelerin 14'ünün farklı ülkelerde cinsiyete dayalı DMF gösterdiği görülmektedir. DMF'nin yönü incelendiğinde ise olguları bilimsel yollarla açıklama yeterliği gerektiren maddelerin çoğunlukla erkekler lehine; veri ve kanıtları bilimsel olarak yorumlama yeterliği gerektiren maddelerin ise çoğunlukla kadınlar lehine DMF gösterdiği görülmektedir. Ülkeler karşılaştırıldığında ise en çok cinsiyete bağlı DMF'li madde içeren ülkeler Türkiye, Hong Kong, İsveç ve İsrail'dir. En az olan ülke ise Fransa'dır. Modele SED değişkeninin eklenmesi Türkiye ve İsveç'te cinsiyete dayalı DMF gösteren madde sayısını azaltırken diğer ülkelerde herhangi bir manidar etki göstermemiştir. Modele okuma değişkeninin eklenmesi ise Hong Kong, Finlandiya ve Türkiye'de cinsiyete dayalı DMF gösteren madde sayısını azaltmıştır. Modele eklenen değişkenlerin Japonya, Birleşik Devletler, Fransa, Brezilya ve İsrail'de cinsiyet DMF'sine herhangi bir etkisi olmamıştır.

SONUÇLAR ve TARTIŞMA

Uluslararası karşılaştırmalarda temel alınan PISA gibi büyük ölçekli sınavlarda bireylerin demografik özelliklerinin test puanlarına etkisinin incelenmesi oldukça önemlidir. Test puanları bireylerin demografik özelliklerinden bağımsız olmalıdır; ancak yapılan birçok çalışmada cinsiyetin büyük ölçekli sınavlar için bir DMF kaynağı olduğu bilinmektedir. Bu bağlamda araştırmanın amaçlarından biri PISA 2015 fen uygulamasında yer alan S12 madde kümesindeki maddelerin dokuz ülkede cinsiyete bağlı DMF gösterip göstermediğinin belirlenmesidir. Araştırma kapsamına dahil edilen ülkelerin tamamında cinsiyete dayalı DMF gösteren maddeler olduğu görülmektedir. Benzer bulgular uluslararası yapılan sınavlara yönelik pek çok çalışmada da raporlanmıştır (Atalay Kabasakal ve Kelecioğlu, 2012; Lan, 2014; Le, 2009; Lyons-Thomas, Sandilands ve Ercikan, 2014).

Araştırma kapsamında ayrıca cinsiyete dayalı DMF gösteren maddelerin özellikleri de incelenmiştir. Araştırmanın sonuçlarına göre Le'nin (2009) bulgularına paralel olarak olguları bilimsel yollarla açıklama yeterliği gerektiren maddelerin çoğunlukla erkekler lehine, veri ve kanıtları bilimsel olarak yorumlama yeterliği gerektiren maddelerin ise çoğunlukla kadınlar lehine DMF göstermektedir. Yalnızca İsrail örneğinde olguları bilimsel yollarla açıklama yeterliği gerektiren bir madde kadınlar lehine, veri ve kanıtları bilimsel olarak yorumlama yeterliği gerektiren bir madde de İsveç ve Fransa örneğinde erkekler lehine DMF göstermiştir. Bu bağlamda madde türünün cinsiyet ile etkileşiminin olabileceği ancak söz konusu etkileşimin farklı kültürlerde farklı şekillerde ortaya çıkabileceği göz önünde bulundurulmalıdır.

Araştırmalarda yalnızca maddelerin gruplar arasında DMF gösterip göstermediğinin belirlenmesi yeterli değildir. Farklı cinsiyetteki öğrenciler için testin yapı geçerliğinin sağlanamamasının altında yatan nedenlerin araştırılması önemlidir. Bu bağlamda araştırmanın ikinci temel amacı SED ve okuma becerisi değişkenlerinin cinsiyete dayalı DMF analizlerinde dolaylı bir etkisinin olup olmadığının belirlenmesidir. SED ve okuma becerisinin başarı ile ilişkisinin yüksek olduğu bilindiğinden araştırmaya dahil edilen maddelerin cinsiyet DMF'sini daha doğru belirleyebilmek amacıyla belirtilen değişkenler modele eklenmiştir. Araştırma kapsamında söz konusu değişkenler her bir ülke için ayrı ayrı değerlendirilmiştir. Söz konusu değişkenlerin modele eklenmesi dört ülkede DMF'li madde sayısını azaltmıştır. Bu bulgular SED ve okuma becerisi değişkenlerinin

cinsiyetle ilişkili DMF'li madde sayısının azaltılması ile ilişkili olduğunu göstermektedir. SED değişkeni "Göktaşları ve Kraterler, Eğim-Yüzey İncelemesi, Beyin Kontrollü Robotik" konularına ilişkin üç maddede Türkiye ve İsveç'te cinsiyete dayalı DMF üzerinde anlamlı bir etkiye neden olmuştur. Okuma becerisi ise "Göktaşları ve Kraterler" ve "Deniz ortamında sesler" konularına ilişkin iki maddede Türkiye, Finlandiya ve Hong Kong'da cinsiyete dayalı DMF üzerinde anlamlı bir etkiye neden olmuştur. Okuma becerisinin modele eklenmesinin kadınlar lehine ortaya çıkmış olan DMF'yi ortadan kaldırmış olması önemli bir bulgu olarak görülebilir. Gelecek çalışmalarda okuma becerisinin cinsiyet DMF'sine etkisi daha kapsamlı bir şekilde değerlendirilerek ortaya çıkan cinsiyet DMF'sinin okuma becerileri arasındaki farktan kaynaklanıp kaynaklanmadığı incelenebilir. Bu çalışmanın sonuçları genel olarak test performansındaki farklılıkların bir kısmının cinsiyetle açıklanabileceğini, ayrıca sosyoekonomik düzey ve okuma becerisinin de bu farklılığa bir katkısı olabileceğini göstermiştir.

DMF, maddelerin sabit bir özelliği değildir. Karşılaştırma yapılan gruplara bağlı olarak değişebilir. Dolayısıyla farklı ülkelerde cinsiyete dayalı DMF'yi etkileyen farklı özelliklerin olması beklenen bir durumdur. Araştırma kapsamına dahil edilen SED değişkeni Türkiye ve İsveç'te; okuma becerisi ise Türkiye, Finlandiya, ve Hong Kong'da cinsiyete dayalı DMF'yi etkilerken diğer ülkelerde her iki değişkenin herhangi bir etkisi gözlenmemiştir. Gelecek çalışmalarda farklı kültürlerde cinsiyete dayalı DMF'yi etkileyebilecek farklı değişkenler incelenebilir. Ayrıca madde türlerinin cinsiyet DMF'sinin yönü ile olan ilişkisinin de kültürlere göre farklılaşabileceği gözlemlenmiştir. Dolayısıyla değişkenlerin etkisinin her bir kültür bağlamında değerlendirilmesi de oldukça önemlidir.

Çalışmada gruplar arası puan karşılaştırılabilirliğini etkileyebileceği düşünülen değişkenler DMF modeline eklenmiştir. Gelecek çalışmalarda öğrenci başarısı ile ilişkili olduğu bilinen ikinci dilde test alma, konu alanı, madde türü gibi diğer dışsal değişkenlerin de gruplar arası karşılaştırmaları nasıl etkileneceği incelenebilir. Ayrıca çalışmada SED ve okuma becerisi sürekli değişkenler olarak ele alınmıştır. Ancak alan yazında genellikle DMF analizlerinde kesikli değişkenler kullanılmaktadır. Bu bağlamda söz konusu değişkenlerin sürekli veya kesikli biçimde ele alınmasının etkisi de incelenebilir.

KAYNAKÇA

- Ackerman, T. A. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. *Journal of Educational Measurement, 29*, 67–91. doi:10.1111/j.1745-3984.1992.tb00368.x
- Allalouf, A., Hambleton, R. K., & Sireci, S. G. (1999). Identifying the causes of DIF in translated verbal items. *Journal of Educational Measurement, 36*(3), 185–198.
- Asil, M. ve Gelbal, S. (2012). PISA öğrenci anketinin kültürler arası eşdeğerliği. *Eğitim ve Bilim, 37*(166), 236-249.
- Atalay Kabasakal, K. ve Kelecioğlu, H. (2012). PISA 2006 öğrenci anketinde yer alan maddelerin değişen madde fonksiyonu açısından incelenmesi. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi, 45*(2), 77-96.
- Barr, A. B. (2015). Family socioeconomic status, family health, and changes in students' math achievement across high school: A mediational model. *Social Science & Medicine, 140*, 27-34.
- Budgell, G. R., Raju, N. S., & Quartetti, D. A. (1995). Analysis of differential item functioning in translated assessment instruments. *Applied Psychological Measurement, 19*(4), 309-321.
- Camilli, G. (1993). The case against item bias techniques based on internal criteria: Do item bias procedures obscure test fairness issues? The use of differential item functioning statistics: A discussion of current practice and future implications. In P. W. Holland, & H. Wainer (Eds.), *Differential item functioning* (pp. 397-413). Hillsdale, NJ: Lawrence Erlbaum.
- Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. London: Sage.
- Candell, G. L., & Drasgow, F. (1988). An iterative procedure for linking metrics and assessing item bias in item response theory. *Applied Psychological Measurement, 12*, 253-260.
- Chun, S. (2014). *Using MIMIC methods to detect and identify sources of DIF among multiple groups*. Unpublished master thesis. University of South Florida, USA.

- Clauser, B., Mazor, K., & Hambleton, R. K. (1993). The effects of purification of the matching criterion on the identification of DIF using the Mantel-Haenszel procedure. *Applied Measurement in Education, 6*, 269-279.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. (1966). *Equality of educational opportunity*. DC: US Department of Health, Education & Welfare. Office of Education (OE-38001 and supp.), 1066-5684.
- Cromley, G. J. (2009). Reading achievement and science proficiency: International comparisons from the programme on international student assessment. *Reading Psychology, 30*(2), 89-118.
- Demps, D. L., & Onwuegbuzie, A. J. (2001). The relationship between eighthgrade reading scores and achievement on the Georgia High School Graduation Test. *Research in the Schools, 8*(2), 1-9.
- Dorans, N. J., & Holland, P.W. (1993). *DIF detection and description: Mantel-Haenszel and standardization*. In P. W. Holland, & H. Wainer (Eds.), *Differential item functioning* (pp. 35-66). Hillsdale: Lawrence Erlbaum Associates.
- Dorans, N. J., & Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement, 23*, 355-368.
- Ercikan, K., & Koh, K. (2005). Examining the construct comparability of the English and French versions of TIMSS. *International Journal of Testing, 5*(1), 23 - 35.
- Fleishman, J. A., Spector, W. D., & Altman, B. M. (2002). Impact of differential item functioning on age and gender differences in functional disability. *Journal of Gerontology: Social Sciences, 57*(5), 275-284.
- Finch, H. (2005). The MIMIC model as a method for detecting DIF: Comparison with Mantel-Haenszel, SIBTEST, and the IRT likelihood ratio. *Applied Psychological Measurement, 29*, 278-295.
- French, B. F., & Maller, S. J. (2007). Iterative purification and effect size use with logistic regression for differential item functioning detection. *Educational and Psychological Measurement, 67*, 373-393.
- Gallo, J. J., Anthony, J. C., & Muthe'n, B. O. (1994). Age differences in the symptoms of depression: A latent trait analysis. *Journal of Gerontology: Psychological Sciences, 49*, 251-264.
- Glöckner-Rist, A., & Hoijtjink, H. (2003). The best of both worlds: Factor analysis of dichotomous data using item response theory and structural equation modeling. *Structural Equation Modeling, 10*, 544-565.
- Grisaya, A., & 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.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
- Hecht, S. A., Burgess, S. R., Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2000). Explaining social class differences in growth of reading skills from beginning kindergarten through fourth-grade: The role of phonological awareness, rate of access, and print knowledge. *Reading and Writing, 12*(1), 99-128.
- Holland, W. P., & Thayer, D. T. (1988). *Differential item performance and the Mantel-Haenszel procedure*. In H. Wainer, & H. I. Braun (Eds.), *Test validity* (pp. 129-145). Hillsdale, NJ: Lawrence Erlbaum.
- Huang, X., Wilson, M., & Wang, L. (2016). Exploring plausible causes of differential item functioning in the PISA science assessment: Language, curriculum or culture. *Educational Psychology, 36*(2), 378-390. <http://dx.doi.org/10.1080/01443410.2014.946890>
- Husin, M. (2014). *Assesing mathematical competence in second language: Exploring DIF evidences from PISA Malaysian data*. Unpublished master thesis, University of Wisconsin, Milwaukee.
- Hyde, J. S., & Lin, M. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104*(1), 53-69. Doi: <http://dx.doi.org/10.1037/0033-2909.104.1.53>
- Joreskog, K., & Goldberger, A. S. (1975). Estimation of a model with a multiple indicators and multiple causes of a single latent variable. *Journal of American Statistical Association, 70*, 631-639.
- Lan, M. C. (2014). *Exploring gender differential item functioning (DIF) in eight grade mathematics items for the United States and Taiwan*. Unpublished doctoral disstertation. University of Washington.
- 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. <http://dx.doi.org/10.1080/15305050902880769>
- Levine, D. W., Bowen, D. J., Kaplan, R. M., Kripke, D. F., Naughton, M. J., & Shumaker, S. A. (2003). Factor structure and measurement invariance of the women's health initiative insomnia rating scale. *Psychological Assessment, 15*, 123-136.
- Logan S., & Johnstone, R. (2009). Gender differences in reading ability and attitudes: examining where these differences lie. *Journal of Research in Reading, 32*(2), 199-214. doi: 10.1111/j.1467-9817.2008.01389.x
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum.

- Lyons-Thomas, J., Sandilands, D. D., & Ercikan, K. (2014). Gender differential item functioning in mathematics in four international jurisdictions. *Education and Science, 39*(172), 20-32.
- MacIntosh, R., & Hashim, S. (2003). Variance estimation for converting MIMIC model parameters to IRT parameters in DIF analysis. *Applied Psychological Measurement, 27*, 372-379.
- Mellenbergh, J. G. (1989). Item bias and item response theory. *International Journal of Educational Research, 13*(2), 127-143.
- Meredith, W., & Millsap, R. (1992). On the misuse of manifest variables in the detection of measurement bias. *Psychometrika, 57*(2), 289-311.
- Millsap, R., & Meredith, W. (1992). Inferential conditions in the statistical detection of measurement bias. *Applied Psychological Measurement, 16*(4), 389-402.
- Muthén, B. O. (1985). A method for studying the homogeneity of test items with respect to other relevant variables. *Journal of Educational Statistics, 10*, 121-132.
- Muthén, B. O., Kao, C. F., & Burstein, L. (1991). Instructionally sensitive psychometrics: Application of a new IRT-based detection technique to mathematics achievement test items. *Journal of Educational Measurement, 28*(1), 1-22.
- Muthén, L. K., & Muthén, B. O. (2010). *Mplus statistical analysis with latent variables user's guide* (6th Edition). Los Angeles, CA: Muthén & Muthén.
- Nolen, S. B. (2003). Learning environment, motivation, and achievement in high school science. *Journal of Research in Science Teaching, 40*(4), 347-368.
- OECD (2015). PISA 2015 technical report. OECD: <http://www.oecd.org/pisa/data/2015-technical-report/>
- Oort, F. J. (1998). Simulation study of item bias detection with restricted factor analysis. *Structural Equation Modeling, 5*, 107-124.
- O'Reilly, T., & McNamara, D. S. (2007). The impact of science knowledge, reading skill, and reading strategy knowledge on more traditional "high-stakes" measures of high school students' science achievement. *American Educational Research Journal, 44*(1), 161-196.
- NCES (2003). NAEP validity studies: An agenda for NEAP validity studies (Report No. 2003-07). Retrieved from <https://nces.ed.gov/pubs2003/200307.pdf>
- Schmidt, W. H., Cogan, L. S., & McKnight, C. C. (2011). Equality of educational opportunity: Myth or reality in U.S. schooling?. *American Educator 34*(4), 12-19.
- Shealy, R. T., & Stout, W. F. (1993). A model-biased standardization approach that separates true bias/DIF from group ability differences and detects test bias/DTF as well as item bias/DIF. *Psychometrika, 58*, 159-194.
- Shih, C. L., & Wang, W. C. (2009). Differential item functioning detection using multiple indicators, multiple causes method with a pure short anchor. *Applied Psychological Measurement, 33*(3), 184-199.
- Sireci, S. G., & Swaminathan, H. (1996). Evaluating translation equivalence: So what's the big DIF? Paper presented at the AERA, Ellenville, NY.
- Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement, 27*, 361-370.
- Thissen, D., Steinberg, L., & Wainer, H. (1988). *Use of item response theory in the study of group differences in trace lines*. In H. Wainer, & H. Braun (Eds.), *Test validity* (pp. 147-169). Hillsdale, NJ: Lawrence Erlbaum.
- Wainer, H., Sireci, S., & Thissen, D. (1991). Differential testlet functioning: Definitions and detection. *Journal of Educational Measurement, 28*, 197-219.
- Walker, C. M., Zhang, B., & Surber, J. (2008) Using a multidimensional differential item functioning framework to determine if reading ability affects student performance in mathematics. *Applied Measurement in Education, 21*(2), 162-181, doi:10.1080/08957340801926201
- Wang, W. C., Shih, C. L., & Yang, C. C. (2009). The MIMIC method with scale purification procedure for detecting differential item functioning. *Educational and Psychological Measurement, 69*(5), 713-731.
- Wang, W. C., & Su, Y. H. (2004a). 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*, 113-144.
- Wang, W. C., & Su, Y. H. (2004b). Factors influencing the Mantel and generalized Mantel-Haenszel methods for the assessment of differential item functioning in polytomous items. *Applied Psychological Measurement, 28*, 450-480.
- Wang, W. C., & Yeh, Y. L. (2003). Effects of anchor item methods on differential item functioning detection with the likelihood ratio test. *Applied Psychological Measurement, 27*, 479-498.
- Welch, C. J., & Miller, T. R. (1995). Assessing differential item functioning in direct writing assessments: Problems and an example. *Journal of Educational Measurement, 32*, 163-178.

- White, K. R. (1982). The relation between socioeconomic status and academic achievement, *Psychological Bulletin*, 91(3), 461-481
- Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago: MESA.
- Woods, C. M. (2009). Evaluation of MIMIC-model methods for DIF testing with comparison to two-group analysis. *Multivariate Behavioral Research*, 44(1), 1-27.
- Woods, C. M., & Grimm, K. J. (2011). Testing for nonuniform differential item functioning with multiple indicator multiple cause models. *Applied Psychological Measurement*, 35(5) 339-361. <http://dx.doi.org/10.1177/0146621611405984>.
- Wu, A. D., & Ercikan, K. (2006). Using multiple-variable matching to identify cultural sources of differential item functioning. *International Journal of Testing*, 6(3), 287-300.
- 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.

EXTENDED ABSTRACT

Introduction

Program of international student assessment (PISA) is an international survey that has been conducted every 3 year by OECD aiming to investigate 15-year-old students' mathematics, science, and reading literacy since 2000. The results of PISA have important implications as it provides governments to compare themselves within an international platform and affect countries' education policies. Hence, concerns about the comparability of PISA results becoming an important and controversial issue.

Measurement instruments are expected to have same psychometric properties for all examinees (Woods, 2009). When an instrument has differentiating psychometric properties between some groups under controlling for differences in the abilities, it is defined as bias. Differential item functioning (DIF) is one of the statistical indicator of bias. Within item response theory perspective, DIF can be defined as differentiation of the probability of correctly answering a question between groups, which have same ability level (Camilli ve Shepard, 1994). When a test has items with DIF, it may reduce the validity of test in group comparisons. Moreover, as the bias causes systematic error, increasing systematic error not only cause decrease in detection of possible differences but also may cause misleading results regarding amount and direction of the difference. Hence, it is important to eliminate DIF items in high stake assessments like PISA.

The purpose of this study was to investigate the presence of DIF between gender groups in PISA 2015 science items in nine selected countries. Moreover, the effect of socioeconomic status and reading ability, on the presence of gender-related DIF were examined, respectively. One cluster from computer-based assessment (CBA) was taken into consideration. The countries were selected among the ones that implemented CBA, on the basis of their rank in science achievement.

Method

The current study was designed as descriptive study since it aims to test validity of PISA 2015 assessment's selected science items. S12 item cluster consisting new items was taken into consideration in the study. S12 cluster consists of 17 items within 5 content area were addressing three competencies: Evaluate and design scientific enquiry, explain phenomena scientifically, interpret data and evidence scientifically, defined in PISA framework.

Population of the study consists of 15-year-old students in Japan, Finland, Hong Kong, United States, France, Sweden, Turkey, Brazil, and Israel. The sample consists of the students who participate PISA 2015 examination and take S12 science cluster in selected countries.

In data analysis, multiple indicators multiple methods (MIMIC) was used. First, items were tested whether they have DIF between gender groups. Then, SES and reading ability included in to the model and their effects on gender DIF was tested. In parameter estimation WLS method was used.

Results and Discussion

The results of the study indicated that 13 items have exhibited gender DIF within different countries. When the interaction between direction of DIF and item competencies examined, in general items requiring explain phenomena scientifically has shown DIF favoring boys and items requiring interpretation of data and evidence scientifically has shown DIF favoring girls. These results are consistent with Le (2009). However, one item in Israel sample and two items in Sweden and France samples contradict the below classification related to item competencies and DIF direction.

The second main objective of the study was to determine whether SES and reading ability have any indirect effects on gender-based DIF analysis. The specified variables were added to the model to better determine the gender DIF on the items specified since they were highly correlated with science achievement. The addition of these variables into the model reduced the number of DIF items in four countries. These findings indicate that the variables of SES and reading ability reduce the number of gender-related DIF items. The SES has a significant impact on gender-based DIF in Turkey and Sweden. The reading ability has a significant impact on gender-based DIF in Turkey, Finland and Hong Kong. In the literature, it is reported that girls generally have higher scores on reading ability than boys (Hyde and Lin, 1988; Logan and Johnstone, 2009). After addition of reading ability into the model, the disappearance of DIF, which was in favor of girls before, can be considered as an important finding.

When the findings are examined in general, it can be seen that 13 of the items in the item cluster reveal gender-based DIF in different countries. When the direction of the DIF is examined, it can be seen that the items, which require the explanation competency of the phenomenon by scientific means, mostly reveal DIF in favor of boys while those require the interpretation competency of data and evidence, mostly reveal DIF in favour of girls. When compared with the countries, the countries having the highest number of items with gender-based DIF are Turkey, Hong Kong, Sweden, and Israel. France has the least number of DIF items. The inclusion of the SES variable into the model reduced the number of gender-based DIF items in Turkey and Sweden, but had no significant effect in other countries. The inclusion of the reading variable into the model, on the other hand, reduced the number of gender-based DIF items in Hong Kong, Finland, and Turkey. The variables added into the model had no effect on gender DIF in Japan, the United States, France, Brazil, and Israel.

It is highly important to examine the effect of demographics of individuals on test scores in large-scale assessments such as PISA, which are considered as base for international comparisons. Test scores should be independent of individual demographics, yet many studies report that gender is a DIF source for large-scale exams. The results of this study, in general terms, show that some of the differences in test performance may be explained by gender. In addition socioeconomic level and reading ability may also contribute to this difference. Further studies can examine how other external variables, such as testing in a foreign language, subject area, item type, which are known to be associated with student achievement, influence intergroup comparisons. In addition, SES and reading ability are considered as continuous variables in the study. In the literature, on the other hand, generally discrete variables are used in DIF analyzes. In this context, the effect of the consideration of the specified variables as continuous or discrete can also be investigated.