

TIMSS 2011 FEN UYGULAMASINDA CİNSİYETE GÖRE FARKLILAŞAN MADDE FONKSİYONUNU AÇIKLAYAN DEĞİŞKENLER

Dr. Seher Yalçın
Dr. Ezel Tavşancıl
Ankara Üniveristesi

Özet

Bu çalışmanın amacı, 2011 yılı Uluslararası Matematik ve Fen Eğilimleri Araştırması (TIMSS) Türkiye 8. sınıf örnekleminde, cinsiyete göre fen maddelerinde farklılaşan madde fonksiyonu (FMF) olup olmadığını, FMF varsa seçilen öğrenci ve okul özelliklerinin FMF gösteren maddelerdeki FMF'nin kaynağını açıklama durumunu çok düzeyli modeller çerçevesinde tespit etmek ve maddelerdeki FMF'nin olası kaynağını yargısal yaklaşımla da belirlemektir. Betimsel araştırma modelinde olan araştırmanın evreni, 2011 yılında Türkiye'de 8. sınıfta öğrenim gören tüm öğrencilerdir. Bu araştırmanın örneklemi, TIMSS 2011 Türkiye uygulamasından oluşmaktadır. TIMSS örneklemi, 2011 yılındaki TIMSS uygulamasına katılması için coğrafi bölgelere ve okul türlerine göre tabakalandırılarak seçilen, 8. sınıf düzeyinde 239 okul ve 6928 öğrenciden oluşmaktadır. Tüm maddeler için FMF, çok düzeyli modelleme yaklaşımında tam Bayes tahminiyle belirlenmiştir. FMF analizleri, tüm kitapçıklar için WinBUGS programı ile yapılmıştır. Araştırma sonucu, FMF gösteren 39 maddenin cinsiyete göre dağılımının eşit olduğu görülmüştür. Öğrenci düzeyinde ele alınan değişkenler iki maddenin FMF gösterme durumunu açıklamıştır. Erkekler lehine FMF olan bu iki madde artık FMF göstermemektedir. Öğrenci düzeyindeki “fene karşı kendine güven” değişkeni, en fazla sayıda maddenin FMF gösterme durumunu açıklamıştır. Okul düzeyinde ele alınan “okul kaynakları” değişkeninin, bazı maddelerin cinsiyete göre FMF gösterme durumu ile ilişkili olduğu görülmüştür. Uzmanlardan alınan görüşler, bazı maddelerde istatistiksel sonuçlarla tutarlıken bazı maddelerde istatistiksel sonuçtan farklı olup bazı maddelerde ise uzmanların kendi aralarında görüş farklılıkları olduğu saptanmıştır.

Anahtar Sözcükler

Çok düzeyli modeller, Fen bilimleri, Farklılaşan madde fonksiyonu, TIMSS 2011.

THE FACTORS EXPLAINING THE DIFFERENTIAL ITEM FUNCTIONING IN THE ADMINISTRATION OF TIMSS 2011 SCIENCE TEST ACCORDING TO GENDER

Dr. Seher Yalcin
Dr. Ezel Tavsancil
Ankara University

Abstract

The objective of this study is to define differential item functioning (DIF) according to the gender and if DIF is determined in the items posd to measure science literacy in TIMSS (Trends in International Mathematics and Science Study) 2011, the role of the students' and schools' characteristics as the sources of DIF in the determined items are analyzed with the frame of multilevel models. Also defining the possible sources of DIF in the items with the judgmental approach is aimed with this study. The whole eight grade students in Turkey are composed of the population of this descriptive study. The sample of this study was consisted of application in TIMSS 2011. TIMSS sampling which it was defined by stratifying according to the geographical regions and the types of schools that attended the TIMSS 2011 and comprised of 239 schools and 6928 eight grade students. In all of the items, DIF is determined through full bayesian estimation in multilevel modeling approach. Analyzes of DIF are progressed by WinBUGS for all booklets. At the end of the study, it is determined that the distribution of 39 items in which DIF is detected is equal by sex. The variables, which are examined in student level, explain DIF in two items. These items displaying DIF in favor of boys don't display DIF anymore. The variable of "self-confidence regarding to science" in student level explains differential item functioning in maximum number of items. It is seen that the variable of "school sources", which is examined in school level, is related with some items displaying DIF by sex. It is determined that while the judgments of experts are consistent with the statistical results in some of the items, they are rather different in another items. Besides, it is detected that there is difference of opinion among experts in some of the items.

Keywords

Differential item functioning, Multilevel models, Science, TIMSS 2011.

GİRİŞ

Teknoloji odaklı gelişmeler her geçen gün hızlı bir şekilde ilerlemekte, bilimsel bilgi birikimi artmakta, bireylerin yaşantılarının her alanında fen ve teknoloji karşlarına çıkmaktadır (Balkan Kıyıcı ve Kıyıcı, 2007; Gömleksiz ve Bulut, 2007; MEB, 2013). Bireylerin yaşadıkları dünyayı anlamaları, fen bilimleri yoluyla gerçekleşir. Bilimsel yöntem ve teknikleri günlük yaşamda karşılaştığı sorunların çözümünde kullanan bireyler, bilgiye daha hızlı ulaşabilir, yeni bilgiler üretebilir, çağdaş teknolojileri etkili ve verimli kullanabilir, yeni sistem ve teknolojiler geliştirebilirler. Bu nedenle fen bilimleri alanı ve etkili olarak öğretilmesi de büyük önem taşır (Kaptan, 1998).

Öğrencilerin hızla gelişen bilimsel ve teknolojik yenilikleri takip edebilmeleri ve anlayabilmeleri için “fen okuryazarı” olmaları gerekmektedir. Fen okuryazarlığı terimi ilk kez Paul Hurd (1958) tarafından kullanılmıştır. Hurd (1958), fen okuryazarlığını, bir bireyin günlük yaşamında, fenle ilgili karşılaşması olası birtakım problemler ve konular hakkında düşünebilmesi ve fenle ilgili kavramların anlamlarını bilmesi olarak tanımlamıştır. Fen ve teknoloji okuryazarı bireylerin yetişebilmesi için öğrencilerin fen alanında sahip oldukları bilgi ve becerilerinin belirlenmesi gerekmektedir.

Alan yazında, pek çok araştırmada, öğrencilerin fen başarılarının cinsiyete göre farklılaştığı tespit edilmiştir (Bursal, 2013; Bursal, Buldur ve Dede, 2015; Fidan-Dişiklitli, 2011). Bu durum Uluslararası düzeyde öğrenci başarılarını belirlemek amacıyla Türkiye’nin de katıldığı araştırmalardan biri olan Uluslararası Matematik ve Fen Eğilimleri Araştırması’nda (TIMSS-Trends in International Mathematics and Science Study) da dikkati çekmektedir. TIMSS 1999 uygulamasında erkek öğrencilerin başarı ortalaması (434), kız öğrencilerden (431) üç puan fazla iken TIMSS 2007 çalışmasında kız öğrencilerin başarı ortalaması (457), erkek öğrencilerin başarı ortalamasından (452) beş puan; TIMSS 2011 de ise kız öğrencilerin ortalama puanı (491), erkek öğrencilerden (475) 16 puan daha yüksektir. Yıllara göre kız ve erkek öğrencilerin fen başarı puanları arasındaki farkın bir miktar arttığı, son iki TIMSS uygulamasında ise kızların puanlarının daha yüksek olduğu görülmektedir. Bu nedenle, öğrencilerin başarıları ve ilgili değişkenlere ilişkin belirlemelerin cinsiyete göre yapılması önem taşımaktadır.

TIMSS uygulamalarında öğrenci başarısının düşük olması, maddelerin herhangi bir alt gruba dezavantaj oluşturulmasıyla ilişkili olabilmektedir. Yapılan ölçme uygulamaları sonucu, ölçülen özellikle ilgili eşit yeteneğe sahip farklı gruplardaki bireylerin yanıtlarının paralel olması yani ölçmelerin farklı gruplar arasında değişmez olması beklenir. Değişmezliğin sağlanamaması puanların yorumunu ve karşılaştırılmasını geçersiz kılmaktadır (Albano ve Rodriguez, 2013).

Ölçme sonuçlarına dayanarak verilen kararların isabetliliği, uygulamaların geçerliği ve güvenilirliği ile yakından ilişkilidir. Kararların geçerliği için mevcut tehditlerden biri madde yanlılığıdır (Clauser ve Mazor, 1998). Yanlılık, ölçme sürecinde sistematik hata olarak tanımlanır (Osterlind, 1983). Bir testteki maddelerin sistematik hata içermesi ise testin geçerliğinin azalmasına neden olmaktadır. Bir testi oluşturan maddelerin yanlı olup olmadığının araştırılması için farklılaşan madde fonksiyonu (FMF) olup olmadığının belirlenmesi gerekir. FMF, aynı yetenek düzeyinde farklı gruplardaki bireylerin bir maddeyi doğru cevaplama olasılığının alt gruplara (odak ve referans) göre farklı olması olarak tanımlanabilir (Embretson ve Reise, 2000; Hambleton, Swaminathan ve Rogers, 1991; Mellenberg, 1989).

TIMSS gibi, sonuçlarına göre ülkelerin eğitim politikalarına ilişkin önemli kararların verildiği uluslararası düzeyde ülkelerin karşılaştırıldığı uygulamalarda yer alan maddelerin FMF içermesi verilen kararların geçerliğini zedelemektedir. Ayrıca bu tür uygulamalarda FMF'nin belirlenmesi kadar aynı zamanda FMF'nin nedenlerinin/ çoklu kaynaklarının belirlenmesi de önemlidir (Albano ve Rodriguez, 2013; Balluerka, Gorostiaga ve Hidalgo, 2010; Beretvas, Cawthon, Lockhart ve Kaye, 2012; Kamata, 2001; Luppescu, 2002; Meulders ve Xie, 2004; Ong, Williams ve Lamprianou, 2011; Turhan, 2006; Williams ve Beretvas, 2006). FMF kaynaklarının belirlenmesi, testin yapı geçerliği tehdidinden sakınılmasını ve yetenek parametre tahminlerinin kesinliğinin artmasını da sağlar (Ong ve diğ., 2011; Turhan, 2006).

FMF'nin belirlenmesinde farklı kuram ve istatistiksel modellere göre farklı yöntemler mevcuttur. Klasik Test ve Madde Tepki Kuramı'na dayalı yöntemlerinin mevcut istatistiksel yapısı FMF'nin olası nedenleri ve yanlılık olup olmadığına ilişkin bilgi verme konusunda oldukça sınırlıdır. Bu tür çalışmalar ancak uzman görüşü alınarak ve alınan uzman görüşü ve açıklanan FMF'li maddeler ile sınırlıdır. TIMSS gibi geniş ölçekli uygulamalarda sınırlı sayıda madde açıklandığından açıklanamayan maddelerde FMF ortaya çıktığında uzman görüşü alınamamakta, maddenin FMF olmasının olası nedeni ve yanlı olup olmadığı hakkında herhangi bir şey ifade edilememektedir. Ancak çok düzeyli ölçme modelleriyle (ÇDOM- Multilevel Measurement Model) yapılan FMF çalışmalarında, maddelerin bilinen özellikleri, FMF'li maddelerde olası FMF kaynaklarını açıklayabilmektedir. Çok düzeyli ölçme modelleri, iç içe geçmiş veri yapılarında önemli bir yer tutmaktadır (Kamata, Bauer ve Miyazaki, 2008).

Bu çalışmada, öğrenci ve okul düzeyinde FMF'yi açıklayabilen değişkenleri belirleyebilmesi, FMF'nin kontrol edilmesine yardım edebilmesi (Binici, 2007; Qian, 2011; Turhan, 2006; Zheng, 2009), bu durumun da yapı geçerliği tehdidinden sakınılmasını sağlaması ve bireyin yeteneğine ilişkin yapılan tahminlerin kesinliğini arttırması (Turhan, 2006) nedenleriyle çok düzeyli FMF belirleme yöntemleri kullanılmıştır. FMF'nin yapı geçerliği için önemli

tehditlerden biri olması, geçerli ve güvenilir ölçmeler yapılabilmesi için FMF'yi açıklayan değişkenlerin belirlenmesini gerekli kılmaktadır.

Türkiye'de fen bilimleri alanında yapılmış ilgili araştırmalara (Ayan, 2011; Bekçi, 2007; Berberoğlu, 1996; Demirtaşlı ve Ulutaş, 2015; Kalaycıoğlu ve Berberoğlu, 2010; Kalaycıoğlu ve Kelecioğlu, 2011; Suna, 2012; Yenal 1995; Yurdugül, 2003) bakıldığında seçme ve yerleştirme çalışmalarında, geniş ölçekli uygulamalarda; çeşitli FMF belirleme yöntemleri ile FMF olup olmadığı belirlenip genellikle yargısal yaklaşımla belirlenen maddelerin yanlı olup olmadığının tespit edildiği görülmektedir. Çalışmalarda, genelde cinsiyete göre FMF olup olmadığı ya da kültürlerarası FMF çalışmaları araştırılmıştır. Yurt dışında ise çok düzeyli modelleme yaklaşımıyla FMF belirleme çalışmalarında (Albano ve Rodriguez, 2013; Atar, 2007; Balluerka ve diğ., 2010; Binici, 2007; Chaimongkol, Huffer ve Kamata, 2007; Cheong, 2006; Kamata ve Binici, 2003; Kamata ve diğ., 2008; Kim, 2003; Luppescu, 2002; Qian, 2011; Vaughn, 2006; Williams, 2003; Williams ve Beretvas, 2006; Zheng, 2009) genellikle gerçek ve simülatif verilerin birlikte kullanıldığı, çok düzeyli modellerin geleneksel yöntemlerle karşılaştırıldığı görülmüş ve çok düzeyli modellerin avantajları vurgulanmıştır.

Fen bilimleri alanının önemi, bu alanda cinsiyete göre öğrenci başarıları arasındaki farklılıklar, FMF'nin kaynağını ortaya çıkarmaya yönelik çalışmaların azlığı ve açıklanan maddeler ile inceleyen uzmanların görüşleriyle sınırlı olması, genelde FMF belirlemede maddeye ilişkin özellikler ve bireylerin az sayıda özelliğinin ele alınması ayrıca ölçmeye konu olan özelliklerin çok düzeyli yapısının ihlal edilerek farklı FMF belirleme yöntemlerinin karşılaştırılmasına odaklanıldığı ve öğrenci ve okul düzeyindeki değişkenlerle az sayıda çalışma yapıldığı görülmektedir. Bu nedenlerle, bu çalışmada sonuçlarına göre eğitim sistemine ilişkin önemli kararların verildiği TIMSS uygulamasının, fen bilimleri alanında, ölçülen özelliğin çok düzeyli yapısına uygun olan, okul ve öğrenci düzeyindeki değişkenleri de bulunduğu düzey içinde analiz edebilen, çok düzeyli FMF belirleme yöntemiyle cinsiyete göre FMF gösteren maddelerdeki değişimi öğrenci ve okul düzeyinde açıklayan değişkenlerin belirlenmesi gereklilik olarak görülmüştür.

YÖNTEM

Çalışma Grubu

Bu araştırma, TIMSS 2011 fen bilimleri uygulaması Türkiye 8. sınıf örnekleminde, cinsiyete göre fen bilimleri maddelerinde FMF olup olmadığını, FMF varsa seçilen öğrenci ve okul özelliklerinin FMF gösteren maddelerdeki FMF'nin kaynağını açıklama durumunu ve FMF'nin olası kaynağını yargısal

yaklaşım ile da belirlemeyi amaçladığından betimsel bir araştırmadır (Freankal, Wallen ve Hyun, 2012).

Evren ve Örneklem

Bu araştırmanın evreni, 2011 yılında Türkiye’de 8. sınıfta öğrenim gören tüm öğrencilerdir. Evrendeki öğrenciler, TIMSS uygulamaları kapsamında, coğrafi bölgelere ve okul türlerine göre tabakalandırılarak iki aşamalı tabakalı örnekleme modeli kullanılarak raslantısal olarak örnekleme seçilmektedir (Joncas, 2008). Bu araştırmanın örneklemi, 2011 yılındaki TIMSS uygulamasına katılan, 8. sınıf düzeyinde 239 okul ve 6928 öğrenciden oluşmaktadır.

Çalışma kapsamında yapılan analizlerde, verinin hiyerarşik olma durumu önem taşıdığından, her okuldan uygulamaya katılan öğrenci sayıları incelenmiştir. Grup büyüklüğü 15’in altında olan okullar analizden çıkarılmıştır (Qian, 2011). Modelin ikinci ve üçüncü düzeyini oluşturan öğrenci ve okul düzeyinde ele alınan değişkenler incelenmiş, ikinci veya üçüncü düzeyde kayıp veriye sahip bireyler analizden çıkarılmıştır. Çıkarılan öğrencilerden sonra her okuldaki öğrenci sayısı 15 ile 52 arasında değişmektedir. TIMSS 2011 sekizinci sınıf fen bilimleri uygulamasında tüm kitapçıklarda toplam 246 madde kullanılmıştır. Bu maddelerden altısı Türkiye’de uygulanmamıştır. Madde düzeyinde kayıp veri oranı %30’dan fazla olan 19 madde, analizden çıkarılmıştır. Analizler 5732 kişi, 208 okul ve 221 madde üzerinden yapılmıştır. TIMSS 2011 uygulamasına katılan öğrenciler 14 farklı kitapçıktan birini almıştır. Kayıp veriler silinmeden önce ve silindikten sonra kitapçık türlerinde öğrencilerin cinsiyete göre dağılımı Çizelge 1’de verilmiştir.

Çizelge 1. TIMSS 2011 uygulamasında kitapçık türlerine göre kayıp veriler silinmeden önceki ve silindikten sonraki cinsiyete göre öğrenci sayıları

Kitapçık Türleri	Silinmeden Önce Öğrenci Sayıları			Silindikten Sonra Öğr. Sayıları		
	Kız	Erkek	Toplam	Kız	Erkek	Toplam
Birinci	270	233	503	232	172	404
İkinci	237	251	488	188	212	400
Üçüncü	239	258	497	204	207	411
Dördüncü	220	277	497	184	222	406
Beşinci	232	258	490	203	202	405
Altıncı	251	243	494	219	200	419
Yedinci	238	256	494	207	206	413
Sekizinci	246	248	494	217	201	418
Dokuzuncu	257	245	502	218	191	409
Onuncu	268	220	488	224	172	396
On birinci	241	255	496	209	210	419
On ikinci	219	267	486	187	213	400
On üçüncü	237	267	504	195	224	419
On dördüncü	259	236	495	220	193	413
Toplam	3414	3514	6928	2907	2825	5732

Çizelge 1’deki kitapçık türlerine göre öğrenci sayıları incelendiğinde, tüm kitapçıklardaki birey sayısının 396 ile 419 değerleri arasında değişmekte ve birbirine yakın olduğu görülmektedir. Farklı kitapçıkları alan öğrencilerin cinsiyete göre dağılımı da birbirine yakındır.

Veri Toplama Araçları ve Kullanılan Materyaller

Çalışmada, TIMSS 2011 uygulaması fen bilimleri başarı testi, öğrenci ve okul anketine ilişkin veriler ve bilgiler IEA’nın web (http://www.iea.nl/timss_2011.html) sayfasından edinilmiştir. TIMSS başarı testlerinde, farklı bilişsel süreçleri ölçmeye yönelik çoktan seçmeli ve yapılandırılmış cevaplı madde türleri bir arada yer almaktadır. Öğrenci düzeyinde analize alınan değişkenler belirlenirken, alan yazından yararlanılarak öğrenci anketindeki maddelerden oluşturulan indeks değişkenler seçilmiştir. Seçilen sekiz değişken; “fene yönelik tutum, fene karşı kendine güven, evdeki çalışma desteği, fen ödevlerine haftalık ayrılan zaman, aile eğitim düzeyi, fen derslerine katılım, fen öğrenmeye verilen değer ve ev eğitim kaynakları”dır. Okul düzeyinde de alan yazından yararlanılarak okul anketinde yer alan maddelerden IEA tarafından oluşturulan “fen öğretiminde okul kaynakları ve akademik başarıya verilen önem” değişkenleri seçilmiştir.

FMF olduğu tespit edilen ve IEA tarafından açıklanan 17 maddenin olası FMF kaynağı konusundaki uzman görüşlerini belirlemek için uzman görüşme formu hazırlanmıştır. Form oluşturulurken istatistiksel analizlerde ele alınan değişkenlerden ve alan yazındaki FMF çalışmalarından yararlanılmıştır (Demirtaşlı ve Ulutaş, 2015; Kalaycıoğlu ve Kelecioğlu, 2011). Hazırlanan formun uygunluğu için ölçme ve değerlendirme alanında iki uzmanın görüşü alınarak gerekli düzeltmeler yapılmıştır. Fen bilgisi öğretmenleri, fen eğitimi ve ölçme ve değerlendirme alanındaki akademisyenlerden oluşan 19 uzmanın görüşüne başvurulmuştur.

Verilerin Çözümlemesi ve Yorumlanması

TIMSS 2011 fen bilimleri uygulaması Türkiye 8. sınıf örnekleminde, cinsiyete göre fen bilimleri maddelerinde FMF olup olmadığı, FMF varsa seçilen öğrenci ve okul özelliklerinin FMF gösteren maddelerdeki FMF’nin kaynağını açıklama durumu Çok Düzeyli Madde Tepki Modellerinden (ÇDMTM) Kamata ve Binici (2003)’nin iki ve üç düzeyli FMF tanımlama modeli ile test edilmiştir. ÇDMTM, verinin iç içe yapısına uygun olarak, grup birimleri içinde yeteneğin değişimine ek olarak okullar gibi grup birimleri arasında da yeteneğin değişimini dikkate alır. Bu nedenle birey düzeyinde yeteneği ve grup düzeyinde yeteneği ayırt etmektedir (Kamata ve Vaughn, 2011).

Veriler analiz edilirken amaca yönelik farklı programlar kullanılmıştır. Verilerin varsayımları karşılama durumu SPSS, STATISTICA, LISREL, R ve BILOG

programları aracılığıyla test edilmiştir. Tüm kitapçıklarda, cinsiyet grupları için ölçme değişmezliği, R yazılım dilinde yer alan “Lavaan” (<http://cran.r-project.org/web/packages/lavaan/index.html>) ve “semTools” (<http://cran.r-project.org/web/packages/semTools/index.html>) paket programları yardımı ile test edilmiştir. FMF analizi, WinBUGS 1.4 (Spiegelhalter ve diğ., 2003) programı ile yapılmıştır.

Analizlere başlamadan önce veri setinden Türkiye’ye ait veriler; sekizinci sınıflar için fen bilimleri testine, öğrenci ve okul anketine verilen cevapları içeren ayrı dosyalar olarak alınmıştır. Okul ve öğrenci düzeyinde seçilen değişkenlerdeki kayıp veriler silinmiştir. Madde düzeyinde; açık uçlu ve çoktan seçmeli maddeler 1-0 olarak yeniden kodlanmıştır. Her bir kitapçıktaki maddeler için kayıp veri yüzdesi incelenmiştir. Kayıp veri yüzdesi %30 ve üzerinde olan maddeler analizden çıkarılmıştır (van Buuren, 2011). Verilerin uç değer içerme durumu incelenmiş herhangi bir uç değere rastlanmamıştır. Tüm maddelerin çarpıklık ve basıklık değerleri $+2$ sınırları içerisinde yer almaktadır. Birinci kitapçık için verilerin MTK’nın varsayımlarından biri olan tek boyutluluk ve yerel bağımsızlık varsayımlarını karşılama durumuna ilişkin incelemeler yapılmış varsayımların sağlandığı görülmüştür.

Araştırma kapsamında karşılaştırmalar cinsiyet grupları arasında yapıldığından, mevcut yapının gruplar arasında değişmez olması gerekmektedir. Ölçme değişmezliği, aynı gerçek puana sahip farklı gruplardaki bireylerin gözlenen puanlarının aynı olması olarak tanımlanabilir (Meredith, 1993; Wu, Li ve Zumbo, 2007). Meredith (1993), ölçme değişmezliğinde Çoklu Grup Doğrulamalı Faktör Analizi’nin (ÇG-DFA) sıkça kullanılan bir yöntemdir. ÇG-DFA, temel model ile model parametrelerinin aşamalı olarak gruplar üzerinde serbest bırakılması veya eşit olacak şekilde sınırlandırılması ile oluşturulan modellerin karşılaştırılmasını içerir. Modeller dört düzeyde ele alınmaktadır i) şekilsel (configural) değişmezlik, ii) zayıf (weak) değişmezlik, iii) güçlü (strong) değişmezlik ve iv) katı (strict) değişmezliktir (Meredith, 1993).

Analizler sonucu, tüm kitapçıklarda şekilsel değişmezliğin sağlandığı, karşılaştırılan gruplar arasında değişmez olduğu görülmüştür. Sadece 11. ve 12. kitapçıklarda değişmezlik testinde ikinci aşama olan zayıf değişmezliğin sağlanmadığı; 7., 8. ve 13. kitapçıklarda ise katı değişmezliğin sağlandığı görülmüştür. Kitapçıklar arasında değişmezliğin sağlanma durumunun farklılaştığı görülmektedir. Tüm kitapçıklarda, iki farklı madde deseninin kullanılması ve kitapçıklar arası hiyerarşik bir desen olduğundan analizler, FMF şüphesiyle tüm kitapçıklar için yapılmıştır.

BULGULAR

Maddelerinin Cinsiyete Göre FMF Gösterme Durumu

İki düzeyli FMF modeline göre yapılan analiz sonucu elde edilen bulgulara göre 221 maddeden B ve C düzeyi FMF genişliğine sahip 39 maddenin 20'si kızlar lehine, 19'u erkekler lehinedir. FMF'li maddelerin kitapçıklara göre dağılımı incelendiğinde; birinci, ikinci, üçüncü, dördüncü, beşinci, altıncı ve onuncu kitapçıklarda 5-7 arası maddede FMF olduğu tespit edilirken onikinci kitapçıkta dokuz, onbirinci kitapçıkta 12 madde de FMF olduğu saptanmıştır. Bu kitapçıklar (11. ve 12. kitapçık), cinsiyete göre ölçülen yapının değişmezliği testinde de sadece şekilsel değişmezliğin sağlandığı kitapçıklardır. Yedinci, dokuzuncu, onüçüncü ve ondördüncü kitapçıklarda 3-4 FMF'li madde tespit edilirken sekizinci kitapçıkta sadece bir maddede FMF saptanmıştır. Yedinci, sekizinci ve onüçüncü kitapçıklar, cinsiyete göre ölçülen yapının değişmezliği testinde de katı değişmezliğin de sağlandığı kitapçıklardır. Sonuçlar bir bütün olarak değerlendirildiğinde, cinsiyete göre ölçme değişmezliğinin sağlanma durumu ile kitapçıklardaki FMF sayısı arasında ters orantı olduğu görülmektedir. Bu beklenen bir durumdur, sonuçların kendi içinde tutarlı olduğunu da göstermektedir.

Öğrenci ve Okul Özelliklerinin Maddelerin Cinsiyete Göre FMF Göstermesini Açıklama Durumu

TIMSS 2011 fen bilimleri uygulamasındaki öğrenci özelliklerinin (fene yönelik tutum, fene karşı kendine güven, evdeki çalışma desteği, ev ödevlerine ayrılan zaman, aile eğitim düzeyi, fen derslerine katılım, fen öğrenmeye verilen değer, ev eğitim kaynakları), maddelerin cinsiyete göre FMF göstermesini açıklama durumunu belirlemek amacıyla yapılan analizler sonucu, öğrenci düzeyinde ele alınan değişkenler, birinci modelde FMF olarak belirlenen iki maddenin (83. ve 186. madde) FMF varyansındaki değişimi açıklamıştır. Bu iki madde artık FMF göstermemektedir. Bu maddelerin ikisi de erkekler lehinedir.

En fazla maddenin FMF gösterme durumunu açıklayan değişken, “fende kendine güven”dir (f: 9). Ardından “ev eğitim kaynakları” (f: 6), “evdeki çalışma desteği” (f: 5) ve “ev ödevlerine ayrılan zaman” (f: 5) değişkenleri gelmektedir. “Fene karşı kendine güven”, “ev eğitim kaynakları”, “fen öğrenmeye verilen değer” değişkenleri erkekler lehine olan sadece bir maddenin FMF göstermesini açıklamaktadır, diğer maddeler kızlar lehinedir. “Ev ödevlerine ayrılan zaman” ve “aile eğitim düzeyi” değişkenleri ise kızlar lehine olan sadece bir maddenin FMF gösterme durumunu açıklamaktadır, diğer maddeler erkekler lehinedir.

TIMSS 2011 fen bilimleri uygulamasındaki FMF'yi açıklayan öğrenci özellikleri kontrol edildiğinde, okul özelliklerinin (fen öğretiminde okul kaynakları ve

akademik başarıya verilen önem), maddelerin cinsiyete göre FMF göstermesini açıklama durumunu belirlemek için yapılan analizler sonucu, 73, 84, 139, 179 ve 218. maddeler için öğrenci özellikleri kontrol edildiğinde, cinsiyete göre FMF'nin kaynaklarından birinin fen öğretiminde okul kaynakları değişkeni olduğu görülmüştür. Bu maddelerden üçü (84, 139 ve 179. maddeler) kızlar lehine, ikisi (73 ve 218. maddeler) erkekler lehinedir.

Cinsiyete Göre FMF Gösteren Maddeler İçin Uzman Görüşleri

TIMSS 2011 fen bilimleri uygulamasındaki cinsiyete göre FMF olduğu tespit edilen ve IEA tarafından açıklanan maddelerin olası FMF kaynağı konusundaki uzman görüşleri sonucu elde edilen bulgulara bakıldığında, uzmanların en fazla FMF'yi açıkladıklarını düşündükleri değişken “fene yönelik tutum”dur. Bu değişkeni, “okulun sahip olduğu kaynaklar” ve “fende kendine güven” değişkenleri takip etmektedir.

Uzmanların görüşleri ile istatistiksel analizler sonucu elde edilen bulgular, bazı maddelerde (16, 32, 35, 40, 73, 81 ve 84) birkaç değişken (aile eğitim düzeyi, fen derslerine katılması, fene yönelik tutum, fende kendine güven, fene verilen değer ve okulun sahip olduğu kaynaklar) için tutarlıyken bazı maddelerde (2, 7, 15, 22, 75, 83, 98 ve 105) tutarlı olmadığı, farklı özelliklere sahip olan uzmanlar arasında da görüş farklılığı olduğu görülmüştür.

TARTIŞMA VE SONUÇ

Araştırma sonucu, çalışma kapsamında incelenen maddelerin yaklaşık %18'inin cinsiyete göre FMF içerdiği görülmüştür. FMF gösteren maddelerin cinsiyete göre dağılımı da yaklaşık olarak eşittir. FMF'li maddelerin kitapçıklara göre dağılımının bir ile 12 arasında değişmesi, farklı kitapçıkları alan öğrencilerin başarılarının kıyaslanmasında geçerli olmayan sonuçlara neden olmaktadır. Aynı zamanda, kitapçıkların birbirine eşit olma durumu hakkında da şüphe uyandırmaktadır. Cinsiyete göre ölçme değişmezliği testi sonucu elde edilen bulgular FMF analizleri sonuçlarıyla birbirini desteklemektedir.

Araştırma kapsamında ele alınan tüm öğrenci düzeyi değişkenler (fene yönelik tutum, fene karşı kendine güven, evdeki çalışma desteği, ev ödevlerine ayrılan zaman, aile eğitim düzeyi, fen derslerine katılım, fen öğrenmeye verilen değer, ev eğitim kaynakları), çeşitli sayıdaki maddenin doğru yanıtlanması ile ilişkilidir. En fazla maddenin doğru yanıtlanma durumu ile ilişkili olan ve en fazla FMF gösteren maddeyi açıklayan değişkenin “fene karşı kendine güven” olduğu saptanmıştır. Bu çalışmanın bulgularına paralel olarak pek çok çalışmada, fen alanında özgüvenin, öğrenci başarıları arasındaki farklılıkları açıklamada manidar istatistiksel etkilerinin olduğu gözlenmiştir (Atar ve Aktan, 2013; Atar ve Atar, 2012; Bayraktar, 2011; Ghagar, Othman ve Mohammadpour, 2011; Kaya ve

Rice, 2010; Kiamanesh, 2004; Thomson ve diğ., 2008; Thomson ve Fleming, 2004; Tighezza, 2014). Birçok ülkede erkeklerle benzer performans göstermelerine rağmen kızların fen alanında benlik kavramı daha düşüktür, erkekler ise daha fazla yüksek öz yeterliliğe sahiptirler (Eurydice, 2010; Mo, Singh ve Chang, 2008; Thomson ve diğ., 2008; Thomson ve Fleming, 2004). Mohammadpour (2012) ise yaptığı çalışmada, kızların fen alanında kendine güveninin erkeklerden daha yüksek olduğunu görmüştür.

FMF gösteren maddelerdeki başarı farklılıklarını açıklama durumu açısından, “fene karşı kendine güven” değişkenini sırasıyla “ev eğitim kaynakları”, “evdeki çalışma desteği” ve “ev ödevlerine ayrılan zaman” değişkenleri takip etmektedir. Bu çalışma sonucu elde edilen bulgulara paralel olarak öğrencilerin sahip olduğu ev eğitim kaynaklarının öğrencilerin fen başarılarıyla manidar ilişkiye sahip olduğu pek çok çalışmada görülmüştür (Abazaoğlu ve diğ., 2014; Acar, 2012; Erberber, 2009; Ghagar, Othman ve Mohammadpour, 2011; Kaya ve Rice, 2010; Kiamanesh, 2004; Thomson ve Fleming, 2004). Öğrencilerin bilgisayar ve internete sahip olmasının da fen bilimleri başarılarında manidar yordayıcılardan biri olduğu bu çalışma sonucu elde edilen bulgulara benzer olarak birçok çalışmada belirtilmektedir (Anıl, 2009; Awang ve Fah, 2013; Bayraktar, 2011; Öztürk ve Uçar, 2010; Thomson, Wernert, Underwood ve Nicholas, 2008).

Öğrencilerin fen başarıları ile ev ödevlerine ayırdıkları zaman arasındaki ilişkilerin incelendiği çalışmalarda, pek çok ülkede öğrencilerin ev ödevlerine ayırdıkları zamanın düşük olmasına rağmen fen başarılarının yüksek olduğu görülmüştür (Acar ve Öğretmen, 2012; Uzun, Bütün ve Yigit, 2010). Thomson ve Fleming (2004), öğrencilerin fazla fen ödevleri yapması ile fen başarıları arasında bir ilişki olmadığını belirtmektedir. Mohammadpour (2012), Rønning (2010) ve Thomson ve diğ. (2008) ise, öğrencilerin fen ödevlerine ayırdıkları zaman arttıkça öğrencilerin fen başarılarının da arttığını ifade etmişlerdir. Yanlılık çalışan bazı çalışmalarda (Gierl, 2005; Wu ve Ercikan, 2007; Yıldırım ve Yıldırım, 2011), öğrencilerin okul dışı ders çalışmaya ayırdığı zaman değişkeninin FMF’nin olası kaynaklarından biri olabileceğini belirtmişlerdir. Briggs (2008) ise her hafta iki saatten fazla ev ödevlerine zaman ayıran öğrencilerin fen başarılarının ayırmayanlardan yüksek olduğunu ifade etmiştir. En az maddenin doğru yanıtlanma durumu ile ilişkili olan ve en az FMF gösteren maddeyi açıklayan değişkenin “fen derslerine katılım” olduğu belirlenmiştir. Fen derslerine katılma değişkeninin fen bilimleri dersindeki akademik başarıyı pozitif yönde etkilediği yapılan çalışmalarda görülmüştür (Chang, Singh ve Mo, 2007; Kahraman, 2014; Mo, 2008; Mo, Singh ve Chang, 2013).

FMF’yi açıklayan öğrenci özellikleri kontrol edildiğinde, okul özelliklerinin (akademik başarıya verilen önem ve okul kaynakları), hiçbir maddenin cinsiyete göre FMF gösterme durumunu tamamen açıklamadığı saptanmıştır. Ayrıca, cinsiyete göre FMF’nin kaynaklarından birinin “okulun sahip olduğu kaynaklar”

olduğu görülmüştür. “okulda akademik başarıya verilen önem”, herhangi bir cinsiyete göre FMF olan madde için manidar bir yordayıcı değildir.

Acar ve Öğretmen (2012) PISA 2006 verileriyle yaptıkları çalışmada, okulda internet bağlantısı bulunan bilgisayar sayısı ve okulun eğitimsel kaynaklarının niteliği arttıkça öğrencilerin fen bilimleri performanslarının da arttığını tespit etmiştir. Oral ve McGivney (2013)’de TIMSS 2011 fen bilimleri alanında yaptıkları çalışmada, ders materyali eksikliği, okul binasındaki ve ısıtma sistemindeki sorunlar ve derslik ortamındaki yetersizliklerin okul için ders anlatımı ile olumsuz ilişki içinde olduğunu tespit etmişlerdir. Büyüköztürk ve diğ. (2014)’nin TIMSS 2011 fen bilimleri alanında yaptıkları çalışmada, okullarında fen ve teknoloji laboratuvarı olan öğrencilerin fen ve teknoloji başarı ortalamalarının, okulunda fen ve teknoloji laboratuvarı olmayan öğrencilerin ortalamasından daha fazla olduğunu ifade etmişlerdir.

Okulda akademik başarıya verilen önem ise, öğrenci düzeyindeki değişkenler kontrol edildiğinde, herhangi bir cinsiyete göre FMF olan madde için manidar bir yordayıcı değildir. Büyüköztürk ve diğ. (2014)’nin TIMSS 2011 fen bilimleri alanında yaptıkları çalışmada ise uygulamaya katılan ülkelerin genelinde, okulun akademik başarıya önem verme derecesi arttıkça öğrencilerin başarılarının artma eğiliminde olduğu görülmüştür. Alan yazındaki “okulun sahip olduğu kaynaklar” değişkeniyle ilgili araştırma sonuçlarının bu çalışmadan elde edilen bulguyla tutarlı olduğu görülmektedir. Okulun akademik başarıya verdiği önem değişkeni için ise alan yazında bu kapsamda tek bir araştırmaya rastlanmış olsa da bulguların tutarlı olmaması, öğrencilerin cinsiyete göre başarı farklılıklarını öğrenci düzeyinde ele alınan değişkenlerin açıklamış olmasından ya da bu çalışma kapsamına dahil edilemeyen değişkenlerle ilişkili olmasından kaynaklanıyor olabilir.

Uzmanların en fazla FMF’yi açıkladıklarını düşündükleri değişken “fene yönelik tutum”dur. Bu değişkeni, “okulun sahip olduğu kaynaklar” ve “fene kendine güven” değişkenleri takip etmektedir. Uzmanların görüşleriyle istatistiksel sonucun en fazla tutarlı olduğu değişken “aile eğitim düzeyi”dir. İstatistiksel sonuçlarla en fazla tutarlı görüş bildiren uzmanların genelde lisans eğitimini fen alanında tamamlamış ve ölçme ve değerlendirme alanında lisansüstü eğitim almış bireyler olduğu görülmüştür. Bu koşullarla birlikte öğretmenlik deneyiminin olması veya yanlılık konusunda çalışmaları olan bireylerin görüşlerinin istatistiksel sonuçlarla daha tutarlı olduğu görülmüştür.

Bulgular genel olarak değerlendirildiğinde, öğrenci ve okul düzeyinde ele alınan özelliklerin bir cinsiyet grubuna avantaj sağlayabildiği görülmüştür. Bu durum, cinsiyete göre başarı farklılıklarının pek çok faktörle etkileşim halinde olduğunu göstermektedir. Ayrıca, öğrencilerin biyolojik farklılıklarının yanında zihinsel farklılıklarının da cinsiyete göre başarı farklılıklarıyla ilişkili olabileceği göz önünde bulundurulmalıdır.

KAYNAKLAR

- Abazaoğlu, İ., Yıldırım, O. ve Yıldızhan, Y. (2014). Türkiye'nin öğretmen profili. *International Periodical For The Languages, Literature and History of Turkish or Turkic*, 9(2), 1-20.
- Acar, T. (2012). Determination of a differential item functioning procedure using the hierarchical generalized linear model: A comparison study with logistic regression and likelihood ratio procedure. *SageOpen*, DOI: 10.1177/2158244012436760, 1-8.
- Acar, T. ve Öğretmen, T. (2012). Çok düzeyli istatistiksel yöntemler ile 2006 PISA fen bilimleri performansının incelenmesi. *Eğitim ve Bilim*, 37(163), 178-189.
- Albano, D. A., and Rodriguez, M. C. (2013). Examining differential math performance by gender and opportunity to learn. *Educational and Psychological Measurement*, 73(5), 836-856.
- Anıl, D. (2009). Uluslararası öğrenci başarılarını değerlendirme programı (PISA)'nda Türkiye'deki öğrencilerin fen bilimleri başarılarını etkileyen faktörler. *Eğitim ve Bilim*, 34, 87-100.
- Atar, B. (2007). *Differential item functioning analyses for mixed response data using IRT likelihood-ratio test, logistic regression, and GLL-AMM procedures*. Unpublished doctoral dissertation, University of Florida State, Tallahassee.
- Atar, B. ve Aktan, D. Ç. (2013). Örtük regresyon iki parametrelili lojistik modeli. *Eğitim ve Bilim*, 38(168), 59-68.
- Atar, H. Y. ve Atar, B. (2012). Türk eğitim reformunun öğrencilerin TIMSS 2007 fen başarılarına etkisinin incelenmesi. *Kuram ve Uygulamada Eğitim Bilimleri*, 12(4), 2621-2636.
- Ayan, C. (2011). *PISA 2009 fen okuryazarlığı alt testinin değişen madde fonksiyonu açısından incelenmesi*. Yayınlanmamış Yüksek Lisans Tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Awang, R., and Fah, L. Y. (2013). An analysis on the selected factors contributing to science and mathematics achievement among secondary students in two SEAMEO member countries. *Jurnal Teknologi (Social Sciences)*, 63(2), 109-116.
- Balkan Kıyıcı, F. ve Kıyıcı, M. (2007). Science, technology & literacy. *The Turkish Online Journal of Educational Technology (TOJET)*, 6(2), 1303-6521.
- Balluerka, N., Gorostiaga, A., Gómez-Benito, J., and Hidalgo, D. (2010). Use of multilevel logistic regression to identify the causes of differential item functioning. *Psicothema*, 22(4), 1018-1025.
- Bayraktar, Ş. (2011). Uluslararası fen ve matematik çalışması (TIMSS 2007) sonuçlarına göre Türkiye'de fen eğitiminin durumu: Fen başarısını etkileyen faktörler. *Selçuk Üniversitesi Ahmet Keleşoğlu Eğitim Fakültesi Dergisi*, 30, 249-270.
- Bekçi, B. (2007). *Orta öğretim kurumları öğrenci seçme ve yerleştirme sınavının değişen madde fonksiyonlarının cinsiyete ve okul türüne göre incelenmesi*. Yüksek Lisans Tezi, Ankara: Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü.

- Beretvas, S. N., Cawthon, S. W., Lockhart, L. L., and Kaye, A. D. (2012). Assessing impact, DIF, and DFF in accommodated item scores: A comparison of multilevel measurement model parameterizations. *Educational and Psychological Measurement*, 72, 754-773.
- Berberoglu, G. (1996). The university entrance examinations in Turkey. *Studies in Educational Evaluation*, 22(4), 363-373.
- Binici, S. (2007). *Random-effect differential item functioning via hierarchical generalized linear model and generalized linear latent mixed model: A comparison of estimation methods*. Unpublished doctoral dissertation, University of Florida State, Tallahassee.
- Briggs, D. C. (2008). Using explanatory item response models to analyze group differences in science achievement. *Applied Measurement In Education*, 21, 89-118.
- Bursal, M. (2013). İlköğretim öğrencilerinin 4-8. sınıf fen akademik başarılarının boylamsal incelenmesi: Sınıf düzeyi ve cinsiyet farklılıkları. *Kuram ve Uygulamada Eğitim Bilimleri*, 13(2), 1141-1156.
- Bursal, M., Buldur, S. ve Dede, Y. (2015). Alt sosyo-ekonomik düzeyli ilköğretim öğrencilerinin 4-8. sınıflar fen ve matematik ders başarıları: Cinsiyet perspektifi. *Eğitim ve Bilim*, 40(179), 133-145.
- Büyüköztürk, Ş., Çakan, M., Tan, Ş. ve Atar, H. Y. (2014). *TIMSS 2011 ulusal matematik ve fen raporu 8. sınıflar*. Ankara: İşkur Matbaacılık.
- Chaimongkol, S., Huffer, F. W., and Kamata, A. (2007). An explanatory differential item functioning (DIF) model by the WinBUG 1.4. *Songklanakarın Journal of Science and Technology*, 29, 449-458.
- Chang, M., Singh, K., and Mo, Y. (2007). Science engagement and science achievement: Longitudinal models using NELS data. *Educational Research and Evaluation*, 13(4), 349-371.
- Cheong, F. Y. (2006). Analysis of school context effects on differential item functioning using hierarchical generalized linear models. *International Journal of Testing*, 6(1), 57-79.
- Clauser, B. E., and Mazor, K. M. (1998). Using statistical procedures to identify differentially functioning test items. *Educational Measurement: Issues and Practice*, 17(1), 31-44.
- Demirtaşlı, N. ve Ulutaş, S. (2015). A study on detecting differential item functioning of PISA 2006 science literacy items in Turkish and American samples. *Eurasian Journal of Educational Research*, 58, 41-60. <http://dx.doi.org/10.14689/ejer.2015.58.3>
- Embretson, S. E., and Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Erberber, E. (2009). *Analyzing Turkey's data from TIMSS 2007 to investigate regional disparities in eighth grade science achievement*. Unpublished doctoral dissertation. Boston College, Boston, Massachusetts.
- Eurydice. (2010). *Eğitim çıktılarında cinsiyet farklılıkları: Avrupa'da alınan tedbirler ve mevcut durum*. Eurydice Raporu, Brüksel. 16 Eylül 2014 tarihinde http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports/120TR.pdf adresinden erişilmiştir.

- Fidan-Dişikiti, A. (2011). *İlköğretim 7. ve 8. sınıf öğrencilerinin fen ve teknoloji dersine yönelik tutumları ile fen ve teknoloji dersi başarıları arasındaki ilişki*. Yayınlanmamış yüksek lisans tezi, Selçuk Üniversitesi Eğitim Bilimleri Enstitüsü, Konya.
- Fraenkel, J. R., Wallen, N. E., and Hyun, H. H. (2012). *How to design and evaluate research in education* (Eighth Edition). New York: McGraw-Hill.
- Ghagar, M. N., Othman, R., and Mohammadpour, E. (2011). Multilevel analysis of achievement in mathematics of Malaysian and Singaporean students. *Journal of Educational Psychology and Counseling*, 2, 285-304.
- Gierl, M. J. (2005). Using dimensionality based DIF analysis to identify and interpret constructs that elicit group differences. *Educational Measurement Issues and Practice*, 24(1), 3-13.
- Gömleksiz, M. ve Bulut, İ. (2007). Yeni fen ve teknoloji dersi öğretim programının uygulamadaki etkililiğinin değerlendirilmesi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 32, 76-88.
- Hambleton, R. K., Swaminathan, H., and Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: SAGE Publications.
- Hurd, P. D. (1958). Scientific literacy: New minds for a changing world. *Issues and Trends*, 82, 407-416.
- Joncas, M. (2008). TIMSS 2007 sampling weights and participation rates. In J. F. Olson, M. O. Martin, & I. V. S. Mullis (Eds.), *TIMSS 2007 technical report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Kahraman, N. (2014). Cross-grade comparison of relationship between students' engagement and TIMSS 2011 science achievement. *Education and Science*, 39(172), 95-107.
- Kalaycıoğlu, D. ve Berberoğlu, G. (2010). Differential item functioning analysis of the science and mathematics items in the university entrance examinations in Turkey. *Journal of Psychoeducational Assessment*, 20, 1-12.
- Kalaycıoğlu, D. B. ve Kelecioğlu, H. (2011). Öğrenci seçme sınavının madde yanlılığı açısından incelenmesi. *Eğitim ve Bilim*, 36, 3-13.
- Kamata, A. (2001). Item analysis by the hierarchical generalized linear model. *Journal of Educational Measurement*, 38 (1), 79-93.
- Kamata, A., Bauer, D. J., and Miyazaki, Y. (2008). Multilevel measurement model. In A. A. O'Connell and D. B. McCoach (Eds.), *Multilevel analysis of educational data* (pp.345-388). Charlotte, NC: Information Age Publishing.
- Kamata, A., and Binici, S. (2003, July). *Random effect DIF analysis via hierarchical generalized linear modeling*. Paper presented at the biannual International Meeting of the Psychometric Society, Sardinia, Italy.
- Kamata, A., Chaimongkol, S., Genc, E., and Bilir, K. (2005, April). *Random-effect differential item functioning across group unites by the hierarchical generalized linear model*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.
- Kamata, A., and Vaughn, B. K. (2011). Multilevel IRT modeling. In J. J. Hox ve J. K., Roberts (Eds.), *Handbook of advanced multilevel analysis* (pp. 41-57). New York: Routledge.

- Kaptan, F. (1998). Fen bilgisi öğretiminin niteliği ve amaçları. Editör: Yaşar, Ş. *Fen bilgisi öğretimi* (s. 13-30). Anadolu Üniversitesi Açıköğretim Fakültesi Yayınları. No: 585, Eskişehir.
- Kaya, S., and Rice, D. C. (2010). Multilevel effects of student and classroom factors on elementary science achievement in five countries. *International Journal of Science Education*, 32(10), 1337-1363.
- Kiamanesh, A. R. (2004, July). *Self-concept, home background, motivation, attribution and their effects on Iranian students' science achievement*. Paper presented at the Third International Biennial SELF Research Conference, Berlin, Germany.
- Kim, W. (2003). *Development of a differential item functioning (DIF) procedure using the hierarchical generalized linear model: A comparison study with logistic regression procedure*. Unpublished doctoral dissertation, University of the Pennsylvania State.
- Luppescu, S. (2002, April). *DIF detection in HLM item analysis*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*, 58, 525-543.
- Meulders, M., and Xie, Y. (2004). Person-by-item predictors. In P. De Boeck and M. Wilson (Eds.), *Explanatory item response models: A generalized linear and nonlinear approach* (pp. 213-240). New York: Springer-Verlag.
- Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research*, 13, 127-143.
- Millî Eğitim Bakanlığı. (2013). *İlköğretim kurumları (ilkokullar ve ortaokullar) fen bilimleri dersi (3, 4, 5, 6, 7 ve 8. Sınıflar) öğretim programı*. Ankara.
- Mo, Y. (2008). *Opportunity to learn, engagement, and science achievement: Evidence from TIMSS 2003 Data*. Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.
- Mo, Y., Singh, K., and Chang, M. (2013). Opportunity to learn and student engagement: A HLM study on eighth grade science achievement. *Educational Research for Policy & Practice*, 12(1), 3-19.
- Mohammadpour, E. (2012). A multilevel study on trends in Malaysian secondary school students' science achievement and associated school and student predictors. *Science Education*, 96(6), 1013-1046.
- Ong, Y. M., Williams, J., and Lamprianou, I. (2011). Exploring differential bundle functioning in mathematics by gender: The effect of hierarchical modelling. *International Journal of Testing*, 11, 271-293.
- Oral, I., and McGivney, E. (2013). *Türkiye'de matematik ve fen bilimleri alanlarında öğrenci performansı ve başarının belirleyicileri: TIMSS 2011 analizi*. İstanbul: Eğitim Reformu Girişimi.
- Osterlind, J. S. (1983). *Test item bias*. London: Sage Publications.
- Öztürk, D. ve Uçar, S. (2010). TIMSS verileri kullanılarak Tayvan ve Türkiye'deki 8. Sınıf öğrencilerinin fen başarısına etki eden faktörlerin belirlenmesi ve karşılaştırılması. *Ç.Ü. Sosyal Bilimler Enstitüsü Dergisi*, 3(29), 241-256.

- Qian, X. (2011). *A multi-level differential item functioning analysis of trends in international mathematics and science study: Potential sources of gender and minority difference among U.S. eighth graders' science achievement*. Unpublished doctoral dissertation, Faculty of the University of Delaware.
- Rønning, M. (2010). *Homework and pupil achievement in Norway: Evidence from TIMSS*. Retrieved from http://www.ssb.no/emner/04/02/20/rapp_201001/rapp_201001.pdf
- Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. (2003). *WinBUGS user manual version 1.4*, MRC Biostatistics Unit, Institute of Public Health and Department of Epidemiology & Public Health, Imperial College School of Medicine. Retrieved from <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/manual14.pdf>.
- Suna, E. (2012). *TIMSS 2007 fen bilimleri testindeki maddelerin dil ve cinsiyet yanlılığı açısından incelenmesi*. Yayımlanmamış yüksek lisans tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Thomson, S., and Fleming, N. (2004). *Examining the evidence: Science achievement in Australian schools in TIMSS 2002* (TIMSS Australia Monograph No 7). Melbourne, VIC: ACER Press.
- Thomson, S., Wernert, N., Underwood, C., and Nicholas, M. (2008). *TIMSS 2007: Taking a closer look at mathematics and science in Australia*. Trends in International Mathematics and Science Study, Camberwell, Vic.
- Tighezza, M. (2014). Modeling relationships among learning, attitude, self-perception, and science achievement for grade 8 Saudi students. *International Journal of Science and Mathematics Education*, 12(4), 721-740.
- Turhan, A. (2006). *Multilevel 2PL item response model vertical equating with the presence of differential item functioning*. Unpublished doctoral dissertation, University of Florida State, Tallahassee.
- Uzun, S., Bütüner, S. Ö. ve Yiğit, N. (2010). 1999-2007 TIMSS fen bilimleri ve matematik sonuçlarının karşılaştırılması: Sınavda en başarılı ilk beş ülke-Türkiye örneği. *Elementary Education Online*, 3, 1174-1188.
- van Buuren, S. (2011). Multiple imputation of multilevel data. In J. J. Hox ve J. K., Roberts (Eds.). *Handbook of advanced multilevel analysis* (pp.173-197). New York: Routledge.
- Vaughn, K. B. (2006). *A hierarchical generalized linear model of random differential item functioning for polytomous items: A bayesian multilevel approach*. Unpublished doctoral dissertation, University of Florida State, Tallahassee.
- Williams, N. J. (2003). *Item and person parameter estimation using hierarchical generalized linear models and polytomous item response theory models*. Unpublished doctoral dissertation, The University of Texas, Austin.
- Williams, N. J., and Beretvas, S. N. (2006). DIF identification using HGLM for polytomous items. *Applied Psychological Measurement*, 30, 22-42.
- Wu, A. D., and Ercikan, K. (2007). Using multiple-variable matching to identify cultural sources of differential item functioning. *International Journal of Testing*, 6(3), 287-300.
- Wu, D. A., Li, Z., and Zumbo, B. D. (2007) Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research & Evaluation*, 12(3), 1-26.

- Yenal, E. (1995). *Differential item functioning analysis of the quantitative ability section of the first stage of the university entrance examination in Turkey*. Yayınlanmamış yüksek lisans tezi, Orta Doğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Yıldırım, H. H. ve Yıldırım, S. (2011). Farklı işleyen madde analizlerinde ortak etken varyansıyla ilişkili eşleme değişkenleri. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 40, 386-396.
- Yurdugül, H. (2003). *Ortaöğretim kurumları seçme ve yerleştirme sınavının madde yanlılığı açısından incelenmesi*. Yayınlanmamış Doktora Tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Zheng, X. (2009). *Multilevel item response modeling: Applications to large-scale assessment of academic achievement*. Unpublished doctoral dissertation, University of California, Berkeley.

EXTENDED ABSTRACT

As differential item functioning (DIF) becomes a threat for structure validity, it is necessary to determine the variables explaining DIF in order to have valid and reliable measurements. The objective of this study is to define DIF according to the gender and if DIF is determined in the items posd to measure science literacy in TIMSS (Trends in International Mathematics and Science Study) 2011, the role of the students' and schools' characteristics as the sources of DIF in the determined items are analyzed with the frame of multilevel models. Also defining the possible sources of DIF in the items with the judgmental approach is aimed with this study. The whole eight grade students in Turkey are composed of the population of this descriptive study. The sample of this study was consisted of application in TIMSS 2011. TIMSS sampling which it was defined by stratifying according to the geographical regions and the types of schools that attended the TIMSS 2011 and comprised of 239 schools and 6928 eight grade students. After deleting the missing values, the analysis were progressed with 208 schools, 5732 students and 221 items. The assumptions of analysis were tested with SPSS, STATISTICA, LISREL, BILOG and R 3.0.3 programs. Analyses of DIF were progressed by WinBUGS 1.4 for all booklets. In all of the items, DIF was determined through full bayesian estimation in multilevel modeling approach. At the end of the study, it was determined that the distribution of 39 items in which DIF was detected was equal by sex. The fact that the number of the items in which DIF was detected per booklets changed from 1 to 12 caused invalid results about comparison of the achievements of students who were administered different booklets. Also, this raised doubt about equality of booklets. According the results of this study, the increase in the values of all variables at student level enables to raise student's achievement. When it comes to results of variables at student level, "self-confidence regarding to science" explained differential item functioning in maximum number of items. Moreover, "self-confidence regarding to science", "home educational resources", "time spent on homework", "home study support", "educational level of family", "attitudes to science", "value of science learning" and "student engagement in science lessons" are related with achievement differences of students according to the gender. The variables "self-confidence regarding to science", "home educational resources", "value of science learning" explained that only one item favoring males showed DIF and the remaining items favor females. The variables "time spent on homework" and "educational level of family" revealed that only one item favoring females showed DIF and the remaining items favor males. The variables, which were examined in student level, explained DIF in two items. These items displaying DIF in favor of boys don't display DIF anymore. It was seen that the variable of "school sources", which was

examined in school level, was related with some items displaying DIF by sex. It can be seen that several factors play role in achievement differences of students according to the gender. It was determined that while the judgments of experts were consistent with the statistical results in some of the items, they were rather different in another item. The variable about which the views of experts and statistical results were more consistent was “education level of family”. Experts whose views were more consistent with statistical results were found to be those who graduated from the field of science and received graduate education about measurement and evaluation. Besides, it was detected that there was difference of opinion among experts in some of the items.

YAZARLAR HAKKINDA

Dr. Seber Yalçın, Ankara Üniversitesi, Eğitim Bilimleri Fakültesi, Ölçme ve Değerlendirme bölümünde araştırma görevlisi olarak görev yapmaktadır.

Prof. Dr. Ezel Taşancıl, Ankara Üniversitesi, Eğitim Bilimleri Fakültesi, Ölçme ve Değerlendirme bölümünden emekli öğretim üyesidir.

İletişim Adresi: Ankara Üniversitesi, Eğitim Bilimleri Fakültesi, Ölçme ve Değerlendirme Bölümü, Cebeci Kampüsü 06590 Ankara. / Eposta: yalcins@ankara.edu.tr.

ABOUT THE AUTHORS

PhD. Seber Yalçın has been working as a research assistant in the department of Measurement and Evaluation in the Faculty of Educational Sciences at Ankara University.

Prof. Dr. Ezel Taşancıl retired lecturer in the department of Measurement and Evaluation in the Faculty of Educational Sciences at Ankara University.

Correspondence Address: Ankara Üniversitesi, Eğitim Bilimleri Fakültesi, Ölçme ve Değerlendirme Bölümü, Cebeci Kampüsü 06590 Ankara, Turkey / Eposta: yalcins@ankara.edu.tr
