

AKÜ FEMÜBİD 19 (2019) 031308 (643-652)

AKU J. Sci. Eng. 19 (2019) 031308 (643-652)

DOI: 10.35414/ akufemubid.615452

Araştırma Makalesi / Research Article

Olasılık Okuryazarlık Testinin Madde Tepki Kuramına Dayalı Yöntemler ile Değişen Madde Fonksiyonu Açısından İncelenmesi

Burcu HASANÇEBİ^{1*}, Yüksel TERZİ², Zafer KÜÇÜK³¹Karadeniz Teknik Üniversitesi, Fen Fakültesi, İstatistik ve Bilgisayar Bilimleri Bölümü, Trabzon.²Ondokuz Mayıs Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü, Samsun.³Karadeniz Teknik Üniversitesi, Fen Fakültesi, İstatistik ve Bilgisayar Bilimleri Bölümü, Trabzon.*Sorumlu yazar e-posta: burcuhsncb@gmail.com ORCID ID: <http://orcid.org/0000-0001-8830-0144>yukselet@omu.edu.tr ORCID ID: <http://orcid.org/0000-0002-6265-542X>zkucuk@ktu.edu.tr ORCID ID: <http://orcid.org/0000-0001-8083-2429>

Geliş Tarihi: 04.09.2019; Kabul Tarihi: 02.12.2019

Öz

Test geliştirme sürecinin birçok adımı vardır. Bu süreçte en önemli adım, testin geçerliliğini ve güvenilirliğini belirlemektir. Madde yanlılığını belirlemek ise, testin geçerliliği ve güvenilirliği için önemli bir kavramdır. Bir maddenin yanlı olması için, Değişen Madde Fonksiyonu içermesi gerekir. Bununla birlikte, yalnızca DMF'nin varlığı, maddeyi yanlı yapmaz. Maddeye verilen cevaplardaki farklılık, alt gruplardaki deneklerin yetenek düzeylerinde farklılık ortaya çıktığında olması gereken bir durumdur. Bu durumu yanlılık olarak yorumlamak doğru değildir. Heterojen bir popülasyon üzerinde bir test yapılacaksa, yanlılık analizi ölçüm sürecinin en önemli bileşeni haline gelir. Buradaki en önemli hedef denekler için adil, hassas ve duyarlı ölçümler yapmaktır. Bu çalışmada, Karadeniz Teknik Üniversitesi İstatistik ve Bilgisayar Bilimleri Bölümü'ndeki 3. ve 4. sınıf öğrencilerinin olasılık teorisi okuryazarlık düzeyleri ölçülmüştür. 3. ve 4. sınıf öğrencilerine 10 maddelik bir okuryazarlık testi uygulanmıştır. Öğrencilerin cevaplarına göre cinsiyet ve sınıf seviyesinde yanlılık analizi yapılmıştır. DMF analizi yaparken, Raju'nun Alan Ölçümü ve Lord'un χ^2 İstatistiği kullanılmıştır. Analizler R yazılımı kullanılarak yapılmıştır. DMF içeren maddeler için uzman görüşüne başvurulmuştur. Analiz sonuçlarına göre, bazı test maddelerinin cinsiyet ve sınıf düzeylerine göre yanlı olduğu tespit edilmiştir.

Anahtar kelimeler

Lord'un χ^2 İstatistiği;

Raju'nun Alan

Ölçümleri;

Madde Yanlılığı;

Değişen Madde

Fonksiyonu.

Investigation of Probability Literacy Test in terms of Differential Item Function by Methods Based on Item Response Theory

Abstract

The process of test development has many steps. The most important step in this process is to determine the validity and reliability of the test. Detection of item bias is an important concept for the validity and reliability of test. In order for an item to be biased, it must contain differential item function (DIF). However, the presence of DIF alone does not make the item biased. When there are differences in the skill levels of the subjects in the sub-groups, it is expected that there will be differences in the responses to the item. Nevertheless, it is not correct to interpret this situation as bias. If a test is to be performed on a heterogeneous population, bias analysis becomes the most important component of the measurement process. The primary objective here is to make fair, precise and sensitive measurement for subjects. In this study, probability theory literacy levels of 3rd and 4th grade students at the Department of Statistics and Computer Sciences of Karadeniz Technical University were measured. A 10-item literacy test was applied to all 3rd and 4th grade students. According to the answers of the students, bias analysis was conducted on gender and class level. When performing DIF analysis, Raju's Area Measurement and Lord's χ^2 Statistics were used with the help of R software. Expert opinion was consulted for items containing DIF. According to the results of the analysis, some test items were found to be biased by gender and grade levels.

Keywords

Lord's χ^2 Statistic;

Raju's Area

Measurement;

Item Bias;

Differential Item

Function.

1. Giriş

Bir ölçme aracının kalitesi güvenilirlik ve geçerlilik çatısı altında toplanır. Güvenilirlik, ölçme aracının hatalardan arınmışlık derecesidir. Ölçmeye muhakkak belirli miktarda hata karışacaktır. Ancak burada önemli olan nokta bu hatanın minimum düzeyde olmasıdır. Bir ölçme aracının güvenilir olması onun hatasız ölçüm yaptığı anlamına gelmez.

Cronbach (1990)'a göre geçerlilik, bir ölçme aracının ölçmeyi amaçladığı özelliği gerçekten ölçüp ölçemediği ile ilgili karar verilebilmesi için kanıt toplama sürecidir. Bu anlamda geçerliliğin ölçme aracının doğruluğu olduğunu söylemek mümkündür (Baykul, 2000). Literatüre bakıldığında sıkça kullanılan geçerlilik kanıtlama yöntemleri, kapsam geçerliliği, ölçüt dayanaklı geçerlilik ve yapı geçerliliğidir.

Madde, bir testin puanlanabilen en küçük birimdir (Turgut 1978). Maddeleri kontrol altında tutarak ve onları amaca hizmet edecek şekilde geliştirerek testi kontrol altında tutmak ve istenilen özelliklere sahip testleri elde etmek mümkündür. Oluşturulan nihai testin bünyesindeki maddelerin istenen özelliklere sahip olup olmadıklarını, eğer sahip değilse neden sahip olmadığının ve sahip olabilmesi için gereken düzeltmelerin bütününe madde analizi denir (Tekin, 1977). Bu süreçte maddelerin niteliği hakkında önemli bilgiler edinilmektedir.

Madde Tepki Kuramı (MTK), kişinin doğrudan gözlemlenemeyen bir özelliği ile geliştirilen testteki maddelere verdiği cevaplar arasındaki ilişkiyi modeller. Bu ilişkiler matematiksel olarak modellenir ve bu modeller MTK modelleri adını alırlar. MTK, ölçme teorisine istatistiksel bir model yaklaşımı sağlaması özelliği ile diğer kuramlardan ayrılmaktadır. MTK'da maddeye verilen cevap iki ya da daha çok kategorili olabilmektedir (Uyar vd.,2013).

MTK modelleri içerdikleri madde parametresi sayısına göre isimlendirilirler. Literatürde bir, iki ve üç parametrelili MTK modelleri mevcuttur. Modellerin barındırdıkları madde parametreleri güçlük, ayırıcılık ve şans parametresidir (Baker, 2016).

Bir parametrelili (1PL) model, MTK modellerinin arasında temel model olarak kabul edilir. Bu model de deneğin maddeye doğru yanıt verme olasılığı sadece kişinin yetenek düzeyi ve madde güçlük indeksine bağlıdır. Şans parametresi sıfır kabul edilir. Ayırıcılık indeksinin ise maddeden maddeye değişmediği düşünülür. Güçlük parametresi maddenin zorluk düzeyinin bir derecesidir (Birnbaum 1968, DeMars 2010, Baker 2016).

İki parametrelili (2PL) MTK modellerinde ise güçlük parametresinin yanında ayırt edicilik indeksi de mevcuttur. Ayırt edicilik indeksi, bir maddenin bilen ile bilmeyeni ayırt etme gücüdür. Bu anlamda maddenin kalitesini gösteren bir parametre olduğu söylenebilir. Bu indeks maddenin geçerlilik ölçüsü olarak da nitelendirilebilir. Bu modelde de şans parametresi sıfır olarak kabul edilir (Birnbaum 1968; DeMars 2010; Baker 2016).

Üç parametrelili (3PL) MTK modellerinde 2PL modele şans parametresi eklenir. Şans parametresi c, düşük yetenek düzeyindeki cevaplayıcıların şans ile maddeye doğru yanıt verme olasılığıdır. c parametresi arttıkça cevaplayıcının yetenek düzeyinin tahminini yapmak zorlaşır ve madde ile ilgili edinilebilecek bilgi miktarı düşer (Birnbaum 1968, DeMars 2010, Baker 2016).

MTK modelleriyle çalışabilmek için üç varsayımın sağlandığı kabul edilir. Bunlar, tek boyutluluk, yerel bağımsızlık ve normallik varsayımlarıdır. Normallik varsayımı ölçülen örtük özelliğin normal dağılım göstermesi şeklinde ifade edilebilir. Bu varsayım günümüzde lojistik regresyon modellerinin kullanımının yaygınlaşması ile birlikte önemini yitirmektedir. Tek boyutluluk, geliştirilen ölçme aracının sadece tek bir özelliği ölçecek şekilde geliştirilmesidir. Yerel bağımsızlık ise farklı test maddelerine verilen yanıtların istatistiksel olarak birbirinden bağımsız olması kabulüne dayanır. Tek boyutluluk varsayımı sağlandığında yerel bağımsızlığında ona bağlı olarak sağlanması beklenir. Literatürde yapılan çalışmalarda geliştirilen ölçme aracının tek boyutlu olması halinde cevaplayıcıların verdikleri yanıtlar arasındaki kovaryansın sıfır olduğu görülmüştür (DeGrujter ve van der Kamp, 2008; Reckase, 2009; DeMars 2010, Baker 2016).

Psikoloji sözlüğüne göre yanlılık, bireyin lehine ya da aleyhine olma eğilimidir (Int Kyn. 1). İstatistiksel açıdan bakıldığında ise, bir kitle parametresi tahmin edildiğinde ortaya çıkan hata miktarıdır (Osterling, 1983). Ölçmedeki yanlılık ise ölçme aracının güvenilir ve geçerli ölçümler yapamaması durumudur. Bir testin yanlı ölçümler yapması belirli bir grup ya da alt testin lehine ya da aleyhine sonuçlar alınmasına yol açar. Bu durumda o ölçme aracının geçerliliği şüpheye düşer. Bu durumu ortadan kaldırmak için ölçme araçlarının yanlılık incelemesi yapılmalıdır. Yanlılık değerlendirmesi madde bazında yapılabileceği gibi testin tamamı için de yapılabilmektedir.

Ölçme aracı içinde yanlı maddeler bulunuyorsa bu maddelerin Değişen Madde Fonksiyonu (DMF) içermesi gerekir. DMF, aynı yetenek düzeyindeki cevaplayıcıların bir madde üzerinde başarılı olma olasılıklarının farklı olması olarak tanımlanabilir (Angoff, 1993). Bir maddenin DMF içerip içermediğine dair istatistiksel analizlerin yapılmasının ardından eğer içeriyorsa DMF içermesine nelerin sebep olmuş olabileceğine dair çikarsamalar yapılmalıdır. DMF kendi başına madde yanlılığının bir kanıtı değildir. DMF belirleme yöntemlerinden elde edilen sonuçlardan yola çıkarak hesaplanan katsayının anlamlı olup olmaması incelenen maddenin DMF gösterip göstermediğini ortaya koyar. İlgilenilen madde DMF içeriyorsa düzeyi ve hangi alt grup için gösterdiği belirlenir. Söz konusu yöntemler kullanılarak farklı gruplarda DMF gösterdiği belirlenen maddeler için uzman görüşü alınarak madde yanlılığına ilişkin karar verilmektedir (Doğan ve Öğretmen 2005, Bakan Kalaycıoğlu ve Kelecioğlu 2011).

DMF belirlenirken cevaplayıcılar ırk, cinsiyet, sosyo-ekonomik durum gibi birçok özelliğe göre alt gruplara ayrılır. Bu gruplar içinde maddenin grup aleyhine çalıştığı düşünülen gruba odak grup, lehine olduğu düşünülen gruba referans grup denir.

DMF'nin belirlenmesi amacıyla çok sayıda istatistiksel yöntem geliştirilmiştir. Bu yöntemler KTK ve MTK'ya dayalı olmak üzere iki ana başlıkta toplanabilir. KTK'ya dayalı yöntemler Mantel-Haenszel, Dönüştürülmüş Madde İndeksi, Lojistik Regresyon, Standartlaştırma yöntemi ve Anova'dır.

MTK'ya dayalı yöntemler ise Lord'un χ^2 İstatistiği, Raju'nun alan ölçümleri ve MTK Olabilirlik Oranı testidir (Atalay v.d. 2012).

Dancer, Anderson ve Derlin (1994) "Use of log-linear models for assessing differential item functioning in a measure of psychological functioning" isimli çalışmada DMF ve madde geçerliliğini hesaplamak için log lineer modelleri kullanmıştır. Araştırmada ruh sağlığı, etnik köken ve cinsiyet değişkenleri kullanılmıştır. Ayrıca araştırmaya 3 etnik gruptan (Afro Amerikan, Anglo Amerikan ve Latin) 627 kadın ve 338 erkek denek katılmıştır. Sonuç olarak, etnik köken ve cinsiyete göre DMF gösteren maddeler bulunmuştur ve ruh sağlığı az olan bireylerde madde cevapları ile cinsiyet ve etnik köken etkileşiminin arttığı ortaya konmuştur.

Allalouf, Hambleton ve Sireci (1999) "Identifying the cause of DIF in translated verbal items" isimli çalışmada farklı dillere çevrilmiş ölçme araçlarında DMF analizi yapmıştır. Ölçme aracı İsrail Psikometrik Giriş Testidir ve İbranice hazırlanmıştır. Rusçaya çevrilip uygulandığı durumdaki DMF analizi yapılmış ve maddelerin %34'ünün diller arasında farklı şekilde çalıştığını göstermiştir.

Gelin ve Zumbo (2003) "DIF results may change depending on how an item is scored: An illustration with the Center for Epidemiological Studies Depression (CES-D) scale" isimli çalışmada DMF sonuçlarının madde puanlama sistemine bağlı olarak oluşup oluşmadığını incelemişlerdir. Araştırmada, depresyon ölçeği kullanılmış ve puanlama sisteminin DMF üzerinde etkisi olduğu sonucuna varılmıştır.

Bakan Kalaycıoğlu ve Kelecioğlu (2011), 2005 yılı ÖSS'de yer alan maddelerin cinsiyete göre DMF içerip içermediği analizini yapmıştır. Test derslere göre dört alt teste bölünmüş DMF değerlerine göre karar verilebilmesi için uzman görüşüne başvurulmuştur.

Atalay Kabasakal, Arsan, Gök ve Kelecioğlu (2014), yaptıkları çalışmada MTK Olabilirlik Oranı, SIBTEST ve Mantel-Haenszel yöntemlerinin karşılaştırmasını ve bu yöntemlerin 1. Tip hata ve güç karşılaştırmalarını yapmışlardır.

Özmen (2014), PISA 2009 okuma becerileri testindeki maddelerin Türkiye ile Amerika ve Birleşik Krallık karşılaştırmaları madde yanlılığı açısından analiz edilmiştir. DMF içerdiği gözlenen maddelerin olası DMF içirme nedenleri belirlenmiştir.

Arıkan Akın (2015), çalışmasında PISA 2012 matematik çalışma etiği tutum maddeleri için değişen madde fonksiyonu analizi yapmıştır. Analiz yapılırken MTK Olabilirlik Oranı, Ordinal Lojistik Regresyon ve Poly-SIBTEST yöntemleri kullanılmıştır. Cinsiyete göre yapılan bu analizlerde 3. maddenin yöntemlere göre farklı düzeylerde olmak üzere DMF gösterdiği sonucuna ulaşılmıştır. Sünbül ve Ömür Sünbül (2016), çalışmasında DMF belirlenmesinde çeşitli yöntemler kullanmışlardır. Ayrıca bu yöntemlerin 1. Tip hata ve güç karşılaştırmaları da analize dahil edilmiştir.

Bu çalışmada, Olasılığa Giriş Okuryazarlık ölçeği geliştirilmiş ve ölçeğin tüm maddeleri için DMF çalışması yapılmıştır. Elde edilen ölçeğin kullanılabilirliği açısından, geçerlilik ve güvenilirlik analizi yapılmıştır. Geçerli ve güvenilir olmayan ölçekler, ölçme çalışmalarında gerçek sonuç vermeyeceğinden kullanılmamalıdır. Geliştirilen bu ölçek ile, Olasılığa Giriş düzeyinde ders alan tüm bölümlerde öğrencilerin seviyeleri hakkında bilgi edinilebilir. Öğrencilerin eksik kaldığı konular belirlenebilir. Bu tip ölçekler, fakültelerin tüm bölümlerinde verilen tüm dersler için uygulanabilir. Literatüre bakıldığında, yapılmış DMF çalışmaları genellikle geliştirilmiş hazır ölçeklere ya da geniş ölçekli kurumsal sınavlara uygulanmıştır. Yapılan çalışma, literatürdeki çalışmaların aksine, boyutun nispeten daha küçük olduğu ve dersin uzmanları tarafından geliştirilmiş bir ölçme aracına uygulanmıştır. Araştırma bu anlamda, tüm fakültelerde verilen tüm dersler için uyarlanabilir.

2. Materyal ve Metot

Tipik bir DMF çalışmasında, bir gruptaki (odak) madde parametreleri ikinci bir grupta (referans) tahmin edilenlerle kıyaslanır. Aynı madde, farklı örneklemelere uygulandığında, DMF'nin bir ölçüsü madde parametrelerinin eşit olup olmadığıdır. Bu eşitliği, MTK parametrelerinin her birinin gruplar arasında özdeş olduğu hipotezini test eden Lord istatistiği kontrol eder. Bu istatistiğin kullanımı için

önemli bir varsayım, yetenek parametresinin (θ) bilinmesidir. (Lord 1980). θ parametresinin ve madde parametrelerinin ikisinin de bilinmemesi halinde 1. Tip hata yapma olasılığı azımsanmayacak miktarda olur (McLaughlin and Drasgow 1987). Lord'un χ^2 istatistiği iki gruptan elde edilmiş madde parametrelerinin arasında fark olmadığı hipotezini test eder.

H_0 : İki grup arasında madde parametreleri açısından fark yoktur ($b_R = b_F$).

H_1 : İki grup arasında madde parametreleri farklılık gösterir ($b_R \neq b_F$).

$$SE(b_R - b_F) = \sqrt{V(b_R) + V(b_F)} \quad (1)$$

$$z = \frac{b_R - b_F}{SE(b_R - b_F)} \quad (2)$$

$$z^2 \sim \chi^2_{(1)}$$

Burada, b_R , referans gruptan elde edilen madde güçlüğü, b_F , odak gruptan elde edilen madde güçlüğü, $V(b_R)$ ve $V(b_F)$ odak ve referans gruptan elde edilen madde güçlüklerinin varyansı ve $SE(b_R - b_F)$ fark vektörünün standart hatasıdır. χ^2 istatistiğinin serbestlik derecesi genellikle test edilen parametre sayısına eşittir.

$$\chi^2_i = v_i' \Sigma^{-1} v_i \quad (3)$$

Burada, v_i , parametreler arası fark vektörü ve Σ varyans kovaryans matrisidir.

3PL model için parametreler arası fark vektörü Eşitlik 4 ile hesaplanır.

$$v'_i = [b_{iR} - b_{iF}, a_{iR} - a_{iF}, c_{iR} - c_{iF}] \quad (4)$$

Parametreler arası fark vektörü v'_i eşitliği ile hesaplanır. Burada, $b_{iR} - b_{iF}$, odak ve referans grupta i . madde için güçlüklerin farkı, $a_{iR} - a_{iF}$, odak ve referans grupta i . madde için ayırt ediciliklerin farkı ve $c_{iR} - c_{iF}$, odak ve referans grupta i . madde için şans parametresinin farkıdır. Eşitlik 4.'de $c_i = 0$ olarak alındığında 2PL modele, a_i parametresi her madde için eşit alındığında 1PL modele indirgenmiş olur.

Lord'un χ^2 istatistiğinin hesaplanması için olabilirlik fonksiyonunun parametre uzayında normal dağılımı ve olabilirlikten türetilen Σ 'nın

tahminlerinin mevcut olması varsayımlarının sağlanması gerekmektedir. Bu hesaplamalarda parametreler arası fark vektörünün kullanılmasının sebebi, hesaplamaların bir elips düzleminde yapılmasıdır. Elipsde x ekseninin genişliği y eksenindeki boyu ile aynı değildir. Dolayısıyla elips üzerindeki noktalar farklı değerlendirilmelidir.

DMF belirlenmesinde iki farklı grubun Madde Karakteristik Eğrisi (MKE) arasında kalan alanın O'dan önemli ölçüde farklı olup olmadığını belirlemeye dayanan yöntem Raju'nun Alan Ölçümleri yöntemidir. Raju 1988 yılındaki çalışmasında 1PL, 2PL ve 3PL için iki MKE arasında kalan işaretli ve işaretli alanların hesaplanması için formülleri ortaya koymuştur. Ayrıca c şans parametreleri eşit olmadığında iki grubun MKE'leri arasında kalan alanın sonsuz olduğunu da göstermiştir. Raju'nun işaretli ve işaretli alan formülleri madde parametrelerinin birer fonksiyonudur. Uygulamada bilinen sadece madde parametrelerinin tahminleri olduğundan bu tahminlere dayanan alanlar örneklemin değişiminden etkilenmektedir. Örneklemin değişmesinden kaynaklanan farkın derecesi hakkında kesin bir fikir yoktur. Ayrıca, iki MKE arasında kalan alanın değişiminin yalnızca örneklem değişiminden mi başka durumlardan etkilenip etkilenmeyeceğini belirlemek de mümkün değildir. Bu durum, iki MKE arasındaki alanı DMF belirlemede kullanan araştırmacılar için önemli bir sorundur (Raju 1990).

$F_1(\theta)$ ve $F_2(\theta)$ aynı madde için grup1 ve grup2'nin 3PL modele göre MKE'leri olsun.

$$F_1 = F_1(\theta) = c_1 + (1 - c_1)P_1 \quad (5)$$

$$F_2 = F_2(\theta) = c_2 + (1 - c_2)P_2 \quad (6)$$

$$P_1 = P_1(\theta) = \frac{\exp [Da_1(\theta - b_1)]}{1 + \exp [Da_1(\theta - b_1)]} \quad (7)$$

$$P_2 = P_2(\theta) = \frac{\exp [Da_2(\theta - b_2)]}{1 + \exp [Da_2(\theta - b_2)]} \quad (8)$$

Burada, a_1 ve a_2 sırasıyla grup1 ve grup2'den elde edilen madde ayırt edicilik indeksleri, b_1 ve b_2 sırasıyla grup1 ve grup2'den elde edilen madde güçlük indeksleri ve D , bir sabittir.

F_1 ve F_2 eşitlikleri hangi MTK modeli kullanılacaksa onun içerdiği parametre sayısına göre $c_i = 0$ alınarak 2PL modele, D göz ardı edilip $a_i = 1$ alınarak 1PL ya da Rasch modele indirgenebilir (Raju, 1990).

Tahmin edilmiş θ 'lar ve madde parametrelerine göre F_1 ve F_2 MKE'ler arasında kalan işaretli ve işaretli alanlar sırasıyla Eşitlik 8 ve Eşitlik 9'da verilmiştir:

$$\begin{aligned} \text{İşaretli Alan} &= SA_{kl} \\ &= \int_{-\infty}^{+\infty} (\widehat{F}_1 - \widehat{F}_2) d\theta \end{aligned} \quad (9)$$

$$\begin{aligned} \text{İşaretsiz Alan} &= UA_{kl} \\ &= \int_{-\infty}^{+\infty} |\widehat{F}_1 - \widehat{F}_2| d\theta \end{aligned} \quad (10)$$

Burada, k madde yanıt teorisi modeli l ise;

$$l = \begin{cases} 0, & a_i \text{ lereşit ya da değil, işaretli} \\ 1, & a_i \text{ lereşit, işaretsiz} \\ 2, & a_i \text{ lereşit değil, işaretsiz} \end{cases}$$

dir. Buradan hareketle, 1PL ya da Rasch Model için Raju alanları;

$$SA_{10} = \widehat{b}_2 - \widehat{b}_1 \quad (11)$$

$$UA_{11} = |\widehat{b}_2 - \widehat{b}_1| \quad (12)$$

2PL Model için Raju alanları;

$$SA = SA_{20} = \widehat{b}_2 - \widehat{b}_1 \quad (13)$$

$$UA_{21} = |\widehat{b}_2 - \widehat{b}_1|; \widehat{a}_1 = \widehat{a}_2 \quad (14)$$

$$UA_{22} =$$

$$\left| \frac{2(\widehat{a}_2 - \widehat{a}_1)}{D\widehat{a}_1\widehat{a}_2} \ln \left\{ 1 + \exp \left[\frac{D\widehat{a}_1\widehat{a}_2(\widehat{b}_2 - \widehat{b}_1)}{\widehat{a}_2 - \widehat{a}_1} \right] \right\} \right|; \quad (15)$$

$$\widehat{a}_1 \neq \widehat{a}_2$$

3PL Model için Raju alanları;

iki MKE arasında kalan alan 3PL model için c parametreleri eşit olmadığında sonsuzdur. Burada; $\widehat{c}_1 = \widehat{c}_2 = c$ durumu göz önüne alınacaktır. Raju (1988)'e göre, işaretli ve işaretli alanlar c

parametrelerinin eŝit olduđu özel durum halinde Raju alanları;

$$SA_{30} = (1 - c)SA_{20} \quad (16)$$

$$UA_{31} = (1 - c)UA_{21}; \widehat{a}_1 = \widehat{a}_2 \quad (17)$$

$$UA_{32} = (1 - c)UA_{22}; \widehat{a}_1 \neq \widehat{a}_2 \quad (18)$$

biçiminde ifade edilir. Burada, a_1 ve a_2 sırasıyla grup1 ve grup2'den elde edilen madde ayırt edicilik indeksleri, b_1 ve b_2 sırasıyla grup1 ve grup2'den elde edilen madde güçlük indeksleri ve c , ŝans parametresidir.

3. Bulgular

Bu alıřmada, Karadeniz Teknik Üniversitesi İstatistik ve Bilgisayar Bilimleri bölümü 3. ve 4. sınıf öğrencilerine uygulanan Olasılık Teorisi Okuryazarlık testinin sonuçlarına göre R yazılımı ile test içerisindeki maddelerin yanlılık içerip içermediğini, içeriyorsa hangi gruba göre (odak ya da referans) yanlılık içerdiği belirlenmeye alıřılmıştır.

Uygulamada test maddelerinin cinsiyet ve sınıf düzeyine göre yanlı olup olmadıkları incelenmiştir. Uygulama 3. ve 4. sınıf öğrencilerinden 104 öğrenciye uygulanmıştır. Söz konusu öğrencilerin 45'i kız 59'u ise erkektir. Ayrıca öğrencilerin 41'i 3. sınıf öğrencisi, 63'ü de 4. sınıf öğrencisidir.

Analiz sonucunda DMF içeren maddelerin bulunması halinde dersin sorumlusundan uzman görüşü alınarak bu maddelerin yanlı olup olmadığına karar verilmiştir. DMF analizi MTK'ya dayalı yöntemlerden Lord'un χ^2 İstatistiđi, Raju'nun alan ölçümleri yöntemleri ile hesaplanmış ve analiz sonucunda bu iki yöntem karşılaştırılmıştır. Test maddeleri dersin kazanımlarından yola çıkarak hazırlanmıştır. Her kazanım için 2 madde oluşturulması uzman görüşüne başvurularak uygun bulunmuştur.

Madde yanlılığı analizlerine başlamadan önce MTK'nın varsayımlarının sağlanıp sağlanmadığı kontrol edilmelidir. İlk varsayım tek boyutluluk varsayımı için Tek Boyutluluk Testi yapılmış ve sonucunda tek boyutluluk kriteri 0.9730 olarak bulunmuştur. Tek boyutluluğun sağlandığının söylenebilmesi için kriterin 1'e oldukça yakın çıkması

gerekmektedir. Tek boyutluluk testinden alınan sonucun 1'e çok yakın olmasından dolayı bu varsayımın sağlandığını söylemek mümkündür (Drasgow and Lissak, 1983).

Diđer bir varsayım yerel bağımsızlık varsayımıdır. Bu varsayımın sağlanması ise tek boyutluluk varsayımının sağlanmasına bağlıdır. Literatüre bakıldığı zaman birçok arařtırmanın tek boyutluluk varsayımının sağlanması halinde yerel bağımsızlık varsayımının buna bađlı olarak sağlanacağını belirttiklerini görmek mümkündür (DeGrujter ve Van Der Kamp, 2008; Recklase, 2009; DeMars 2010, Baker 2016).

Son varsayım ise normallik varsayımıdır. Normallik varsayımı ölçme teorisinin ilk yıllarında, Normal Ogive Model'in ya da Klasik Test Kuramı'nın (KTK) kullanıldığı dönemlerde, önemlilik arz eden bir varsayımıdır. Ancak günümüzde MTK modelleri yani lojistik regresyon modelleri kullanılmaya başlandıkça normallik varsayımı önemini kaybetmiştir.

Varsayımların sağlanmasıyla MTK modellerinin kullanılabilirliđi ortaya çıkmıştır. Buna bađlı olarak da MTK'ya dayalı DMF yöntemlerinden Lord'un χ^2 İstatistiđi yöntemi ile cinsiyet açısından yanlılık hesaplamalarından elde edilen p deđerleri Çizelge 1.'de verilmiştir.

Çizelge 1. Lord'un χ^2 İstatistiđine Dayalı Referans Grup Kadın Olduđunda Cinsiyet Açısından Yanlılık.

	İstatistik	p-Deđer
1	1.4845	0.2231
2	0.3660	0.5452
3	0.3659	0.5453
4	2.0681	0.1504
5	0.2809	0.5961
6	2.1519	0.1424
7	0.0047	0.9455
8	0.9955	0.3184
9	0.0003	0.9866
10	0.2872	0.5920

Anlam düzeyi $\alpha = 0.05$ olmak üzere, hesaplanan p deđerleri örneklem farklılıklarından etkilenmektedir. Bu etkiyi ortadan kaldırmak için etki büyüklüğü deđerleri hesaplanmış ve bu hesaplamalara göre her bir madde için yanlılık derecesinin deđerleri sunulmuştur. Etki büyüklüğü kodları ise Çizelge 2.'de verilmiştir.

Çizelge 2. DeltaLord ve DeltaRaju Değerlerine göre Yanlılık Derecelendirilmesi.

DeltaLord veya DeltaRaju Değerleri	Yanlılık Derecesi	
$0 < dL < 1$ veya $0 < dR < 1$	A	DMF yoktur
$1 < dL < 1.5$ veya $1 < dR < 1.5$	B	Orta düzeyde DMF mevcuttur
$1.5 < dL $ veya $1.5 < dR $	C	Yüksek düzeyde DMF mevcuttur

Çizelge 2.'deki etki büyüklükleri değerleri baz alınarak örneklem etkisinden arındırılmış DeltaLord değerleri Çizelge 3.'de verildiği gibidir. Çizelge 3.'deki yanlılık derecelerine bakarak Lord'un χ^2 İstatistiğine dayalı referans grup kadın olduğunda cinsiyet açısından 4. ve 6. maddelerin C düzeyinde, 1. ve 8. maddelerin B düzeyinde DMF içerdikleri görülmektedir. Bu durumda ilgili maddelerin kadınların lehine çalıştıklarını söylemek mümkündür.

Çizelge 3. Lord'un χ^2 İstatistiğine Dayalı Referans Grup Kadın Olduğunda Cinsiyet Açısından Yanlılık için Etki Büyüklüğü Değerleri.

	DeltaLord	
1	1.4342	B
2	-0.6850	A
3	-0.6848	A
4	1.6441	C
5	0.5946	A
6	-1.7510	C
7	0.0761	A
8	-1.3035	B
9	0.0197	A
10	0.6557	A

Madde yanlılığının analiz edildiği bir diğer özellik öğrencilerin sınıf düzeyidir. 3. ve 4. Sınıflar arasında yapılan yanlılık analizinden elde edilen p değerleri Çizelge 4.'de verilmiştir.

Çizelge 4. Lord'un χ^2 İstatistiğine Dayalı Referans Grup 3. Sınıf Olduğunda Sınıf Düzeyi Açısından Yanlılık.

	İstatistik	p-Değeri
1	0.0001	0.9939
2	2.9440	0.0862
3	0.8189	0.3655
4	1.4664	0.2259
5	0.3614	0.5477
6	0.5312	0.4661
7	0.0012	0.9729
8	2.8708	0.0902
9	0.3432	0.5580
10	0.2091	0.6475

Çizelge 4.'de verilmiş p değerlerini örneklem etkisinden arındırmak için etki büyüklükleri

hesaplanmıştır. Hesaplanan etki büyüklükleri DeltaLord değerleri ile Çizelge 5.'de verilmiştir. Buna göre Lord'un χ^2 İstatistiğine dayalı referans grup 3. sınıf olduğunda sınıf düzeyi açısından 2. ve 8. Maddelerin C düzeyinde, 3. ve 4. maddelerin B düzeyinde DMF içerdikleri görülmektedir yani ilgili maddeler 3. sınıfların lehine çalışmaktadır.

Çizelge 5. Lord'un χ^2 İstatistiğine Dayalı Referans Grup 3. Sınıf Olduğunda Sınıf Düzeyi Açısından Yanlılık.

	DeltaLord	
1	-0.0094	A
2	-1.9550	C
3	-1.0375	B
4	-1.3940	B
5	-0.6874	A
6	0.9266	A
7	-0.0385	A
8	2.8447	C
9	0.7464	A
10	0.6047	A

MTK'ya dayalı bir diğer yöntem olan Raju'nun Alan Ölçümleri yöntemi için cinsiyet açısından yanlılık analizinden elde edilen p değerleri Çizelge 6.'da verilmiştir.

Çizelge 6. Raju'nun Alan Ölçümlerine Dayalı Referans Grup Kadın Olduğunda Cinsiyet Açısından Yanlılık.

	İstatistik	p-Değeri
1	-1.2184	0.2231
2	0.6050	0.5452
3	0.6049	0.5453
4	-1.4381	0.1504
5	-0.5300	0.5961
6	1.4669	0.1424
7	-0.0683	0.9455
8	0.9977	0.3184
9	-0.0168	0.9866
10	-0.5359	0.5920

Çizelge 6.'da verilmiş p değerlerinin örneklem etkisinden kurtulmaları için hesaplanan etki büyüklüğü değerleri yani DeltaRaju değerleri hesaplanmış ve Çizelge 7.'de verilmiştir. Bu değerler baz alınarak Raju'nun Alan Ölçümlerine dayalı referans grup kadın olduğunda cinsiyet açısından 4. ve 6. maddelerin C düzeyinde, 1. ve 8. maddelerin B düzeyinde DMF içerdikleri yani ilgili maddelerin kadınların lehine olduğu görülmektedir.

Çizelge 7. Raju'nun Alan Ölçümlerine Dayalı Referans Grup Kadın Olduğunda Cinsiyet Açısından Yanlılık için Etki Büyüklüğü Değerleri.

DeltaRaju		
1	1.4342	B
2	-0.6850	A
3	-0.6848	A
4	1.6441	C
5	0.5946	A
6	-1.7510	C
7	0.0761	A
8	-1.3035	B
9	0.0197	A
10	0.6557	A

Sınıf düzeylerine göre DMF analizi için referans grup 3. Sınıflar olmak üzere Raju'nun alan ölçümleri yöntemiyle hesaplanan p değerleri Çizelge 8.'de verilmiştir.

Çizelge 8. Raju'nun Alan Ölçümlerine Dayalı Referans Grup 3. Sınıf Olduğunda Sınıf Düzeyi Açısından Yanlılık.

	İstatistik	p-Değeri
1	0.0077	0.9939
2	1.7158	0.0862
3	0.9049	0.3655
4	1.2110	0.2259
5	0.6011	0.5477
6	-0.7288	0.4661
7	0.0340	0.9729
8	-1.6943	0.0902
9	-0.5859	0.5580
10	-0.4572	0.6475

Raju'nun alan ölçümleri yöntemiyle hesaplanmış ve Çizelge 8.'de verilmiş sınıf düzeyleri açısından yanlılık analizinin p değerlerinin örneklem farklılıklarından kaynaklanan değişimlerden arındırılması için hesaplanan etki büyüklüğü değerleri DeltaRaju sonuçları Çizelge 9.'da verildiği gibidir. Bu değerlerden hareketle referans grup 3. Sınıflar olduğunda 2. ve 8. Maddeler C düzeyinde, 3. ve 4. Maddelerin B düzeyinde DMF içermektedir yani ilgili maddeler 3. Sınıfların lehine çalışmaktadır.

Çizelge 9. Raju'nun Alan Ölçümlerine Dayalı Referans Grup 3. Sınıf Olduğunda Sınıf Düzeyi Açısından Yanlılık için Etki Büyüklüğü Değerleri.

DeltaRaju		
1	-0.0094	A
2	-1.9550	C
3	-1.0375	B
4	-1.3940	B
5	-0.6874	A
6	0.9266	A
7	-0.0385	A

8	2.8447	C
9	0.7464	A
10	0.6047	A

4. Tartışma ve Sonuç

Analizlerin sonucunda aynı referans gruplar için Lord'un χ^2 İstatistiğine ve Raju'nun Alan Ölçümlerine bakıldığında iki yöntemin aynı maddeleri yanlı olarak işaret ettikleri görülmektedir. Buradan hareketle, DMF'nin belirlenmesinde MTK'ya dayalı yöntemlerden Lord'un χ^2 İstatistiği ve Raju'nun Alan Ölçümleri yöntemlerinin birbirlerine üstünlükleri bulunmadığını söylemek mümkündür. Bu alandaki literatür incelendiğinde de yöntemlerin benzer sonuçlar verdiği görülmüştür. Cohen ve Kim (1993), çalışmasında yeterli örneklem büyüklüklerinde ve yeterince uzun testlerde bu iki yöntemin hangisinin seçileceğinin çok önemli olmayacağını her iki yöntem ile benzer sonuçların elde edileceğini ortaya koymuştur. Raju, Drasgow ve Slinde (1993) ise çalışmasında Lord'un χ^2 İstatistiği ile Raju'nun işaretli ve işaretsiz alanlar yöntemlerinin benzer sonuçlar verdiğini ortaya koymuştur.

Bu çalışmada ise dersin uzmanları tarafından geliştirilen Olasılığa Giriş Okuryazarlık Testi isimli ölçme aracı bünyesindeki maddelerde hem cinsiyet açısından hem de sınıf düzeyleri açısından DMF varlığı incelenmiştir. Elde edilen sonuçlara göre her iki yöntem için de cinsiyet açısından 1., 4., 6., ve 8. maddelerde, sınıf düzeyleri açısından 2., 3., 4. ve 8. maddelerde DMF varlığı saptanmıştır. DMF içerdiği belirlenen maddelerin seçilmiş referans grupların lehine çalıştığını söylemek mümkündür. Burada, seçilmiş referans gruplar cinsiyet için kadın grubu, sınıf düzeyleri için ise 3. sınıflardır.

Bir maddenin yanlı olabilmesi için tek başına DMF içermesi yeterli değildir. Analizler sonucu DMF içerdiği belirlenen maddeler için dersin uzmanları tarafından uzman görüşü alınmalıdır. Bu çalışmada, DMF içerdiği gözlenen 1, 2, 3, 4, 6 ve 8. maddeler için dersin uzmanları tarafından uzman görüşü alınmıştır. Uzman görüşüne göre, DMF içerdiği belirlenen maddelerden 1., 4., 6. ve 8. maddelerin

kadınlar lehine; 2., 3., 4. ve 8. maddelerin ise 3. sınıflar lehine yanlı oldukları belirlenmiştir.

Dikkat edilmesi gereken bir nokta da literatürde MTK'ya dayalı yöntemlerden bilinen iki yöntem olan Lord'un χ^2 İstatistiği ve Raju'nun Alan Ölçümleri yöntemlerinin her iki referans grupta (kadınlar ve 3. sınıflar) ve her iki açıdan (cinsiyet ve sınıf düzeyi) aynı sonuçları vermesidir. Bundan dolayı, yeterli örneklem büyüklüğü ve yeterince madde içeren ölçme araçları söz konusu olduğunda MTK'ya dayalı bu iki yöntemin de kullanılabilceğı söylenebilir.

Daha ileriki çalışmalar için, geliştirilmiş bu ölçme aracındaki maddelerin DMF içerip içermedikleri KTK ve MTK'ya dayalı yöntemler ile karşılaştırılabilir. Her iki kurama dayalı yöntemlerden hangisinin daha tutarlı ve etkin sonuçlar verdiğini tartışılabilir. Ayrıca, literatürde bulunan hazır ölçekler ile de yanlılık çalışmaları yapıp yöntemlerin güçleri karşılaştırılabilir.

5. Kaynaklar

- Allalouf, A., Hambleton, R.K. and Sireci, S.G., 1999. Identifying the Causes of DIF in Translated Verbal Items. *Journal of Educational Measurement*, **36(3)**, 185-198.
- Angoff, W. H., 1993. Differential Item Functioning, Paul W. Holland and Howard Wainer (Eds.), NJ:Lawrence Erlbaum Associates, Publishers, 3-23.
- Arıkan Akın, Ç., 2015. Değişen Madde Fonksiyonunu Belirlemede MTK-Olabilirlik Oranı, Ordinal Lojistik Regresyon ve Poly-Sıbtest Yöntemlerinin Karşılaştırılması. *E-Uluslararası Eğitim Araştırmaları Dergisi*, **6(1)**, 1-16.
- Atalay Kabasakal, K., Arsan, N., Gök, B. ve Kelecioğlu, H., 2014. Değişen madde fonksiyonunun belirlenmesinde MTK olabirlik oranı, SIBTEST ve Mantel-Haenszel yöntemlerinin performanslarının (I. tip hata ve güç) karşılaştırılması. *Kuram ve Uygulamada Eğitim Bilimleri*, **14(6)**, 2175-2193.
- Atalay, K., Gök, B., Kelecioğlu, H. ve Arsan, N., 2012. Değişen Madde Fonksiyonunun Belirlenmesinde Kullanılan Farklı Yöntemlerin Karşılaştırılması: Bir Simülasyon Çalışması. *Hacettepe Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, **43**, 270-281.

- Bakan Kalaycioğlu, D. ve Kelecioğlu, H., 2011. Öğrenci Seçme Sınavı'nın madde yanlılığı açısından incelenmesi. *Eğitim ve Bilim*, **36(161)**, 161, 4-13.
- Baker, F.B., 2016. Madde Tepki Kuramının Temelleri. Prof. Dr. Neşe Güler (çeviri editörü), Pegem Akademi, 17-40.
- Baykul, Y., 2000. Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması. Ankara: ÖSYM Yayınları, 193-212.
- Birnbaum, A., 1968. Some Latent Trait Models and Their Use in Inferring an Examinee's Ability. Frederic M. Lord and M.R. Novick (Eds.), Adison-Wesley, 397-479.
- Cohen, A.S., Kim, S., 1993. A comparison of Lord's χ^2 and Raju's area in detection of DIF. *Applied Psychological Measurement*, **17(1)**, 39-52.
- Cronbach, L. J., 1990. Essentials of psychological testing, New York: Harper Collins Publisher, 475.
- Dancer, L.S., Anderson, A.J. and Derlin, R.L., 1994. Use of log-linear models for assessing differential item functioning in a measure of psychological functioning, *Journal of Consulting and Clinical Psychology*, **62(4)**, 710-717.
- De Gruijter, DNM ve van der Kamp, LJT., 2008. Statistical Test Theory for the Behavioral Sciences, *Psychometrika*, **73(4)**, 791-792.
- DeMars, C., 2010. Item Response Theory: Understanding Statistics Measurement. Prof. Dr. Hülya Kelecioğlu (çeviri editörü), Oxford University Press, Oxford, 3-31.
- Doğan, N. ve Öğretmen, T., 2005. Test ve madde yanlılığı. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, **5(1)**, 89-103.
- Drasgow, F. and Lissak, R., 1983. Modified parallel analysis: a procedure for examining the latent dimensionality of dichotomously scored item responses. *Journal of Applied Psychology*, **68(3)**, 363-373.
- Gelin, M. N. and Zumbo, B. D., 2003. Differential Item Functioning Results May Change Depending on How An Item is Scored: An Illustration with the Center for Epidemiologic Studies Depression (CES-D) Scale. *Educational and Psychological Measurement*, **63(1)**, 65-74.
- Lord, F.M. 1980. Applications of Item Response Theory to Practical Testing Problems, Routledge, 52-60.
- McLaughlin, M.E. and Drasgow, F. 1987. Lord's chi-square test of item bias with estimated and with known person parameters. *Applied Psychological Measurement*, **11(2)**, 161-173.

- Özmen, D.T., 2014. PISA 2009 Okuma Testi Maddelerinin Yanlılıđı Üzerine Bir alıřma. *Eđitim Bilimleri ve Uygulama*, **13(26)**, 147-165.
- Raju, M.S., Drasgow, F. and Slinde, J.A. 1993. An emprical comparison of the area methods, Lord's χ^2 test and the Mantel-Haenszel technique for assessing differential item functioning. *Educational and Psychological Measurement*, **53(2)**, 301-314.
- Raju, N.S., 1988, The area between two item characteristic curves, *Psychometrika*. **53(4)**, 495-502.
- Raju, N.S., 1990. Determining the significance of estimated signed and unsigned areas between two item response functions. *Applied Psychological Measurement*. **14(2)**, 197-207.
- Reckase, M.D., 2009. Multidimensional Item Response Theory. New York: Springer, 11-54.
- Sünböl, Ö. ve Ömür Sünböl, S., 2016. Deđişen Madde Fonksiyonunun Belirlenmesinde Kullanılan Yöntemlerde I. Tip Hata ve Güç alıřması, *İlköđretim Online*, **15(3)**, 882-897.
- Tekin, H., 1997. Eđitimde Ölçme ve Deđerlendirme. Ankara: Mars Matbaası, 243-249.
- Turgut, M.F., 1978. Eđitimde Ölçme ve Deđerlendirme. Ankara: Nüve Matbaası.
- Uyar, ř., Öztürk Gübeř, N. ve Keleciođlu H., 2013. PISA 2009 Tutum Anketi Madde Puanlarının Ařamalı Tepki Modeli ile İncelenmesi, *Eđitim ve Öđretim Arařtırmaları Dergisi*, **2(4)**, 125-134.

İnternet kaynakları

1-<http://www.psikolojisoszlugu.com/>, (13.03.2018)