



Değişen Madde Fonksiyonunun Belirlenmesinde “difR” R Paketinin Kullanımı: Ortaöğretime Geçiş Sınavı Fen Alt Testi

MAKALE TÜRÜ	Başvuru Tarihi	Kabul Tarihi	Yayın Tarihi
Araştırma Makalesi	04.02.2020	23.09.2020	26.09.2020

Betül Alathı ¹ ve Selma Şenel ²
Balıkesir Üniversitesi

Öz

Değişen Madde Fonksiyonunun (DMF) belirlenmesi, bir testin ve testten alınan puanların geçerliğine ilişkin önemli göstergeler sunmaktadır. difR paketi ise farklı DMF belirleme yöntemlerinin uygulanmasına izin vererek araştırmacılara ve uygulayıcılara büyük kolaylık sağlayan R paketidir. Bu araştırmanın temel amacı örnek bir araştırma verisi üzerinden, difR paketinde farklı DMF belirleme yöntemlerine ilişkin; yazılım kurulumu, varsayımların incelenmesi, analiz adımları ve analiz sonuçlarının yorumlanması için izlenen sürecin resmedilmesidir. Bu temel amaç doğrultusunda Türkiye’de 8. sınıf öğrencilerine uygulanan Ortaöğretime Geçiş Sınavı 2018 uygulamasında yer alan fen maddelerinin, madde sıra etkisi bakımından DMF gösterme durumları incelenmiştir. Bu yönüyle araştırma tarama modelinde bir araştırmadır. Araştırmada sıklıkla kullanılan DMF belirleme yöntemlerinden Klasik Test Kuramına dayalı Mantel-Haenszel, Lojistik Regresyon ve SIBTEST ile Madde Tepki Kuramına dayalı Olabilirlik Oran yöntemlerine ilişkin adımlar ele alınmıştır. DMF analizleri sonucu elde edilen bulgulara göre fen maddelerinin madde sıra etkisi bakımından çoğunlukla DMF göstermediği ya da ihmal edilebilir düzeyde DMF gösterdiği sonucuna ulaşılmıştır.

Anahtar sözcükler: difR, ölçme değişmezliği, değişen madde fonksiyonu, madde sıra etkisi, Ortaöğretime Geçiş Sınavı.

Etik Kurul Kararı: Bu araştırma, 01.01.2020 tarihinden önce yapıldığı için etik kurul kararı zorunluluğu taşımamaktadır.

¹Sorumlu yazar: Dr. Öğr. Üyesi, Necatibey Eğitim Fakültesi, Eğitim Bilimleri Bölümü, e-posta: betulkarakocalatli@gmail.com, <https://orcid.org/0000-0003-2424-5937>

²Dr. Öğr. Üyesi, Necatibey Eğitim Fakültesi, Eğitim Bilimleri Bölümü, e-posta: selmahocuk@gmail.com, <https://orcid.org/0000-0002-5803-0793>

Eğitimde ve psikolojide, bireylerin zeka, başarı, ilgi tutum gibi birçok özellik, ilgili özelliğe yönelik geliştirilmiş testler ile gözlenebilir hale gelebilmektedir. Testlerden alınan puanların belirli ölçütlerle karşılaştırılması sonucu ise bireylerin test ile ölçülen özelliğinden yola çıkarak bireylerle ilgili önemli kararlar verilmektedir. Seçme, yerleştirme, sınıflama, not verme, durum belirleme gibi farklı biçimlerde olabilen bu kararlar, birçok durumda bireyin yaşamını önemli ölçüde etkileyebilmektedir. Ölçme araçları ve ölçme araçlarından elde edilen sonuçlara ilişkin kararların adil ve yerinde olabilmesi için ölçme araçlarının ölçme değişmezliği gösterdiğine ilişkin kanıtlar ortaya konulması önerilmektedir. Ölçme değişmezliği, belirli bir gizil yapıyı ölçmek üzere geliştirilen testlerden ilgili yapıya ilişkin aynı düzeye sahip olan bireylerin, alt ölçekler ve maddeler düzeyinde aynı gözlenen puanlara sahip olma durumları olarak tanımlanmaktadır (American Educational Research Association-AERA, American Psychological Association-APA ve National Council on Measurement in Education-NCME, 1999; Crocker ve Algina, 1986; International Test Commission-ITC, 2005).

Alanyazın incelendiğinde eğitimde ve psikolojide kullanılan ölçme araçlarının ölçme değişmezliğini incelemek üzere test düzeyinde Yapısal Eşitlik Modellemeleri (YEM), madde ve alt testler düzeyinde ise Değişen Madde Fonksiyonu (DMF) ve Değişen Test Fonksiyonları (DTF) yaklaşımlarının kullanıldığı görülmektedir (Cheung ve Rensvold, 2000; Raju, Laffitte ve Byrne, 2002; Stark, Chernyshenko ve Drasgow, 2006; Vandenberg ve Lance, 2000). DMF, farklı alt gruplarda yer alan ancak aynı yetenek düzeyinde olan bireylerin testte yer alan bir maddeye ilişkin başarı veya doğru yanıt olma olasılıklarının farklılaşması olarak tanımlanmaktadır (Camilli ve Shepard, 1994; Zumbo, 1999). DMF, tüm yetenek dağılımı boyunca sabit bir şekilde bir grup lehine avantaj sağlıyorsa tek biçimli DMF olarak tanımlanmaktadır. Belirli bir yetenek düzeyine kadar bir gruba, belirli bir yetenek düzeyinden sonra diğer grup lehine avantaj sağlıyorsa tek biçimli olmayan DMF olarak adlandırılmaktadır (Swaminathan ve Rogers, 1990).

Alanyazında farklı matematiksel süreçleri izleyen, farklı kesme noktaları ve algoritmalar kullanan çok sayıda DMF belirleme tekniği kullanılmaktadır. Bu yöntemler temel olarak, Klasik Test Kuramı (KTK) ve Madde Tepki Kuramına (MTK) dayalı olmak üzere sınıflandırılmaktadır. Maddelerin ölçme değişmezliğinin incelenmesinde birden fazla DMF belirleme tekniğinin bir arada kullanılması ve elde edilen sonuçların mutlaka karşılaştırılarak karar verilmesi önerilmektedir. Ancak yöntemlerin her birinin farklı istatistiksel yazılımlarda uygulama olanağı olabildiği görülmektedir. Bu durumun bir sonucu olarak DMF'nin incelendiği araştırmalarda, birden fazla istatistik yazılımının bir arada kullanıldığı görülmektedir (Adedoyin, 2010; Akalın, 2014; Çepni, 2011; Gök, Kelecioğlu ve Doğan, 2010; Grover ve Ercikan, 2017; Karakoç-Alatlı ve Çokluk-Bökeoğlu, 2018; Lyons-Thomas, Sandilands ve Ercikan, 2014; Stoneberg, 2004; Walzebug, 2014; Yıldırım, 2015). Örneğin, Koyuncu, Aksu ve Kelecioğlu (2018) tarafından yapılan çalışmada farklı yazılımları karşılaştırmak üzere altı farklı yazılım kullanılmıştır. Bunlar DIFAS, JMETRIK, EZDIF, SPSS (ZumboSyntax), IRTPRO ve IRTLRFIDIF yazılımları olarak

belirtilmiştir. Araştırmada, Uluslararası Öğrenci Değerlendirme Programı (PISA) 2012 uygulaması öğrenci anketi matematik çalışma disiplini maddelerinin matematik performansı bakımından alt ve üst gruplara göre DMF gösterip göstermediği Mantel-Haenszel (MH), Lojistik Regresyon (LR) ve MTK'ye dayalı Olabilirlik Oran (MTK-OO) yöntemleri ile incelenmiştir. DMF çalışmaları için kullanılan yazılımlar ve elde edilen sonuçların karşılaştırılması araştırma konuları arasında yerini almıştır. Tablo 1'de yaygın kullanılan DMF belirleme yöntemleri, bağlı oldukları kuram, tek biçimli ya da tek biçimli olmayan DMF belirleme olanakları ve yöntemde karşılaştırılabilen grup sayısı ile analize olanak tanıyan yazılımlar Tablo 1'de sunulmuştur.

Tablo 1

Değişen Madde Fonksiyonu Belirleme Yöntemleri

Kuram	DMF Belirleme Yöntemi	Grup Sayısı	Tek biçimli/ Tek biçimli olmayan	Yazılım
KTK	Breslow-Day ki-kare	2	Tek biçimli olmayan	DIFAS,
	Mantel-Haenszel	2	Tek biçimli	DIFAS, JMETRİK, EZDIF, Easy-DIF
	Simultaneous Item Bias Test-SIBTEST	2	Tek biçimli	SIBTEST
	Standardizasyon	2	Tek biçimli	Easy-DIF
	Dönüştürülmüş Madde Güçlükleri	2	Tek biçimli	MS Excel
	Lojistik regresyon	2	Her ikisi	EZDIF, SPSS (ZumboSyntax)
	Genelleştirilmiş lojistik regresyon	>2	Her ikisi	SAS, STATISTICA
	Genelleştirilmiş Mantel-Haenszel	>2	Tek biçimli	GMHDIF
MTK	Olabilirlik oranı testi	2	Her ikisi	IRTLRDIF, IRTDIF, ConQuest
	Lord'un ki-kare testi	2	Her ikisi	IRTLRDIF, IRTDIF
	Raju'nun alanı	2	Her ikisi	IRTLRDIF, IRTDIF
	Genelleştirilmiş Lord'un ki-kare testi	>2	Her ikisi	IRTLRDIF, IRTDIF

Tablo 1 incelendiğinde, DMF incelemeleri için araştırmacıların birden fazla yazılımda işlem yapması gerekmektedir. Bu noktada, tüm bu yöntemlerin analiz edilebilmesine ve sonuçların karşılaştırmalarına olanak sağlayan bir yazılım olan R açık kaynak kodlu yazılımı dikkat çekmektedir. R yazılımı kullanıcılar tarafından istatistiksel paketler geliştirilmesine ve diğer kullanıcılarla paylaşılmasına olanak sağlaması, programlama dilinin güçlü olması ve ücretsiz olarak ulaşılabilmesi gibi üstünlüklere sahiptir (Beaujean, 2013). Bu nedenlerle R'nin kullanımı da son yıllarda

giderek yaygınlaşmaktadır. R'nin sınırlılığı ise kullanıcı dostu bir ara yüze sahip olmaması ve bu nedenle diğer istatistik programlarına göre öğrenilmesinin daha zaman alıcı olması olarak görülmektedir (Doğan ve Uluman, 2016). Örneğin, ölçek geliştirme ve ölçme değişmezliği incelemeleri psikolojik ölçme aracı geliştirilen tüm alanlarda yapılabilir. Ancak bu analizlerin R'de olduğu gibi betik dili gerektirmesi, ağırlıklı olarak yalnızca ölçme ve istatistik gibi alanlardan araştırmacıların ölçme değişmezliği araştırmaları yapmasına neden olabilmektedir. Farklı alanlardaki araştırmacıların da ölçme değişmezliği çalışmalarında önemli bir yere sahip olan DMF analizlerini kolaylıkla yapılabilmesi önemlidir. R'nin bu dezavantajlarını ortadan kaldırmak adına yazılımın kullanımı sırasında kullanılan paket programlar, kodlar, analiz çıktıları ve yorumlarının açık bir şekilde ele alındığı çalışmalara gereksinim duyulmaktadır. Buna ilişkin olarak R yazılımının kullanımına ilişkin Er ve Sönmez (2005) R yazılımı kullanımının istatistik eğitiminde önemine vurgu yaparak yazılıma ilişkin genel bilgiler ve temel istatistiklere ilişkin bir çalışma yapmıştır. Horgan (2012) tarafından yapılan bir çalışmada R yazılım dilinin temel özellikleri ele alınmıştır. Doğan ve Uluman (2016) yaptıkları çalışmada R'yi genel anlamda tanıtmış ve eğitim bilimleri alanında, SSCI dergilerde yayınlanmış çalışmalarda kullanılan paket programların dağılımını incelemiştir.

Magis, Beland, Tuerlinckx ve Boeck (2010) ise ikili puanlanan veriler için DMF belirleme yöntemlerine ilişkin R yazılımındaki parametreleri genel olarak özetlemiştir. Örnek veri seti üzerinden MH yöntemine ilişkin kod ve çıktı dosyasının sunulduğu çalışmada altı DMF tekniği için de analiz sonuçlarının karşılaştırıldığı kod ve analiz çıktısına da yer verilmiştir. Araştırma sonucunda, difR paketinin diğer DMF belirleme yazılımlarına göre üstünlükleri ve büyük veri setinde analiz hızının düşük olması gibi sınırlılıkları tartışılmıştır. Stiglic, Watson ve Cilar (2019) hemşirelik araştırmalarında genellikle ticari istatistik paket programlarının kullanıldığı, ücretsiz, güçlü ve esnek bir alternatif olan R yazılımı ise daha az kullanılmaktadır. Buna göre daha önceden bir çalışma için toplanmış veriler üzerinden puan hesaplama, korelasyon grafiği ve betimsel istatistiklerin elde edilmesi gibi temel analiz adımları ele alınmıştır. Alanyazın incelendiğinde R yazılımı farklı analizler üzerinden işlem adımları ve analiz çıktılarının yorumu bakımından çalışma konusu olarak yer almaktadır. Bu araştırma ile R yazılımının kurulumundan, difR paketi ile MH, LR, MTK-OO ve SIBTEST DMF belirleme yöntemlerine ilişkin işlem adımları, kodlarıyla birlikte sırasıyla ele alınmış, analiz çıktıları ise yorumlarıyla birlikte sunulmuştur. DMF inceleme çalışmalarında MTK'ya dayalı yöntemlerin ve bununla birlikte birden çok DMF belirleme tekniğinin bir arada kullanılmasının önerildiği bilinmektedir (Hambleton, 2006). Her bir DMF belirleme yönteminde farklı eşitleme ölçütleri, DMF'li olarak kabul etmek için farklı kesme noktaları ve algoritmaların kullanıldığı ve genel olarak aralarında tam bir uyum olmadığı görülmektedir (Çepni, 2011; Gök, ve diğ., 2010). Bu nedenle bu çalışmada dört farklı DMF tekniği ele alınmıştır. R yazılımında DMF incelemeleri için yol gösteren bu türlü bir çalışmanın yapılmasının araştırmacılar ve uygulayıcılar için oldukça yararlı olacağı değerlendirilmiştir.

Mantel Haenszel Yöntemi

İkili puanlanan maddeler için uygun bir yöntemdir. Odak ve referans grupta yer alan bireylerin testten aldıkları puanlar dikkate alınarak aynı puanı alan bireyler, aynı yetenek düzeyi olarak adlandırılır. Her bir yetenek düzeyi için odds diğer bir deyişle olabirlik oran değerleri hesaplanır. Odds bir olayın gerçekleşme olasılığının gerçekleşmeme olasılığına oranı olarak adlandırılmaktadır. Referans ve odak grup için her bir yetenek düzeyinde doğru ve yanlış yanıtlayan kişi sayısının maddeyi yanıtlayan toplam kişi sayıları dikkate alınarak hesaplanan odds değerleri hesaplanmaktadır (Mertler ve Vannatta, 2005). Elde edilen bu oran 0 ile ∞ arasında değer alır. Camilli ve Shepard'ın (1994) önerisiyle odds değerlerinin yorumlanmasını kolaylaştırmak amacıyla doğal logaritması, -2.35 katı alınarak Δ_{MH} değeri elde edilmektedir. Elde edilen Δ_{MH} istatistiği DMF derecesini ortaya koyan ve DMF etki büyüklüğü olarak adlandırılan bir değerdir. Maddelere ait Δ_{MH} değerinin negatif olması referans grup için, pozitif olması ise odak grup için bir üstünlük sağladığı anlamını taşımaktadır. Δ_{MH} değerinin aldığı mutlak değere göre DMF dereceleri belirlenebilmektedir (Dorans ve Holland, 1993).

- $|\Delta_{MH}| < 1$ DMF ise yoktur veya göz ardı edilebilecek düzeyde (A düzeyi),
- $1 < |\Delta_{MH}| < 1.5$ ise orta düzeyde (B düzeyi),
- $|\Delta_{MH}| \geq 1.5$ ise yüksek düzeyde DMF göstermektedir (C düzeyi).

Lojistik Regresyon Yöntemi

MH yönteminin pratik olması ve sıklıkla kullanılmasının yanı sıra I. tip hata olasılığı yüksek bir yöntemdir. Bununla birlikte MH'nin yalnızca tek biçimli DMF'yi belirlemek üzere kullanılması bu yöntemin sınırlılıkları olarak ortaya çıkmaktadır. Bu noktada hem tek biçimli hem de tek biçimli olmayan DMF'yi belirlemek üzere kullanılan LR yönteminin daha üstün olduğu söylenebilir. Bu yöntemde maddeye verilen yanıt bağımlı değişken; toplam puan, grup değişkeni ise bağımsız değişken olarak tanımlanmaktadır. DMF belirlemek amacıyla bağımsız değişkenler hiyerarşik olarak modele dahil edilirler. Model 1'de toplam puan, Model 2'de grup değişkeni (odak ve referans), Model 3'te ise toplam puan ve grup değişkeni birlikte modele dahil edilir. Modellerden elde edilen R^2 değerleri karşılaştırılarak DMF'nin varlığına ve düzeyine karar verilir. Jodoin ve Gierl'a (2001) göre modeller arası R^2 farkı ile belirlenen ΔR^2 değerine göre DMF düzeyleri aşağıdaki şekilde önerilmektedir:

- $0 < \Delta R^2 < 0.035$ ise DMF yoktur ya da ihmal edilebilir düzeyde
- $0.0035 \leq \Delta R^2 < 0.07$ ise orta düzeyde,
- $\Delta R^2 \geq 0.07$ ise yüksek düzeyde DMF göstermektedir.

Zumbo ve Thomas'ın (1996) ΔR^2 aralıkları da DMF düzeyini değerlendirmede sıklıkla kullanılmaktadır.

- $\Delta R^2 < 0.13$ ise DMF yoktur veya göz ardı edilebilecek düzeyde,
- $0.13 \leq \Delta R^2 < 0.26$ ise orta düzeyde,
- $\Delta R^2 \geq 0.26$ ise yüksek düzeyde DMF göstermektedir.

Madde Tepki Kuramına Dayalı Olabilirlik-Oran Yöntemi

MTK'ya dayalı Olabilirlik Oran yöntemi odak ve referans grup için tüm maddelere ilişkin madde parametrelerinin eşit olduğu sınırlandırılmış model ile i. maddeye ilişkin madde parametresinin eşit olmadığı ancak i. madde dışındaki diğer maddelere ait madde parametrelerinin eşit olduğu genişletilmiş model olmak üzere iki model kurulmaktadır. İki model için elde edilen olabilirlik değerlerinin logaritmaları alınarak aralarındaki farka dayalı olarak G^2 değeri hesaplanır. G^2 değeri modelin parametre sayısı kadar serbestlik derecesinde χ^2 (ki-kare) dağılımı göstermektedir. G^2 değeri χ^2 tablo değeri ile karşılaştırılır. $G^2 > \chi^2_{(tablo\ değeri)}$ ise ilgili madde DMF göstermektedir. Maddenin DMF gösterdiği belirlendiğinde, sırasıyla diğer madde parametreleri için analize devam edilir (Thissen, 2001). Greer (2004) tarafından G^2 değerinin aldığı değerlere göre DMF düzeyleri aşağıdaki gibi özetlenmektedir:

- 3.84 < G^2 < 9.4 ise DMF yoktur veya göz ardı edilebilecek düzeyde,
- 9.4 ≤ G^2 < 41.9 ise orta düzeyde,
- G^2 ≥ 41.9 ise yüksek düzeyde DMF göstermektedir.

SIBTEST Yöntemi

Shealy ve Stout (1993) tarafından geliştirilen, parametrik olmayan ve KTK'ya dayalı bir DMF belirleme yöntemidir. Bu yöntem için, gözlenen puana dayanan MH ve benzeri diğer yöntemlere göre referans ve odak gruplarını eşitleyen çok daha doğru bir eşitleme yaptığı belirtilmektedir (Osterlind ve Everson, 2009). SIBTEST yönteminde test, DMF'nin incelendiği madde ile testte yer alan diğer maddeler olmak üzere ikiye ayrılır. En önemli özelliği DMF incelenen madde dışındaki maddelerden alınan gözlenen toplam puanlar üzerinden, regresyona dayalı yaptığı bir düzeltme ile gerçek puan kestirimleri yaparak odak ve referans grup bireylerini bu puanlara göre eşleştirmesidir. Bu düzeltme ile I. tip hatanın kontrol altına alınması ve elde edilen sonuçların da bu hatalardan daha az etkilenmesi amaçlanmaktadır. Buna göre bu yöntemin istatistiksel gücünün yüksek ve küçük örneklemelerin bulunduğu durumlarda da etkili olduğu belirtilmektedir (Bolt, 2000; Gierl, Jodoin ve Ackerman, 2000). Analiz sonucu elde edilen β indeksi etki büyüklüğünü göstermektedir. Bu indeksin pozitif olması maddenin referans grup lehine, negatif olarak belirlenmesi ise odak grup lehine DMF'nin gözlemlendiği biçiminde yorumlanmaktadır. Bununla birlikte β indeksine göre DMF düzeyi de belirlenebilmektedir (Gotzmann, Wright ve Rodden, 2006; Stout ve Roussos, 1995); Buna göre,

- $|\beta| < 0.059$ ise göz ardı edilebilecek düzeyde,
- $0.059 \leq |\beta| < 0.088$ ise orta düzeyde,
- $|\beta| \geq 0.088$ ise yüksek düzeyde DMF içermektedir.

Değişen Madde Fonksiyonu incelemelerinde maddenin belirli bir değişken bakımından gruplara göre farklı fonksiyon gösterip göstermediğine ilişkin gözlemler yapılmaktadır. Buna göre DMF çalışmaları incelendiğinde cinsiyet (Bakan-Kalaycıoğlu ve Kelecioğlu, 2011; Ünsal-Özberk ve Koç, 2017); kültür ve dil (Asil ve Gelbal, 2012; Karakoç-Alatlı ve Çokluk-Bökeoğlu, 2018); sosyo-ekonomik düzey

(Ünsal-Özberk ve Koç, 2017) yerleşim yeri, okul türü (Osadebe ve Agbure, 2018), engel durumu (Bolt ve Ysseldyke, 2008) gibi değişkenlerin ele alındığı görülmektedir. DMF çalışmalarında konu olan bir diğer durum da aynı testin madde sıraları farklı olan formlarını alma durumunun yanlılığa neden olup olmamasıdır.

Test güvenliğini arttırmak amacıyla maddelerin çeşitli şekillerde sıralandığı farklı kitapçıklar elde edilmektedir. Bu yöntem, öğrencilerin kopya çekme gibi testten alınan puanların güvenilirliğini olumsuz yönde etkileyebilecek durumlara yönelik bir önlem olarak düşünülmektedir. Bir testte yer alan maddelerin sıralanışları bakımından farklılaşan test formlarını yanıtlayan bireylerin, ilgili testten alınan puanları farklılaşabilmekte, test ve madde istatistikleri de bu durumdan etkilenebilmektedir (Balta ve Ömür-Sünbül, 2017; Barciovski ve Olsen, 1975; Çokluk, Gül ve Doğan-Gül, 2016; Doğan-Gül ve Çokluk Bökeoğlu, 2018; Hambleton ve Traub, 1974; Klinker, 1980; Tan, 2009). Bu etkiye madde "sıra etkisi" adı verilmektedir (Kingston ve Dorans, 1984; Yen, 1980). Testin farklı formları geliştirilirken madde sıra etkisinin dikkate alınması testin geçerliliği için de önemli bir husus olarak görülmektedir (Hahne, 2008). Bir testin, maddelerin sıralanışları bakımından farklı formlarını alan gruplara göre ölçme değişmezliği gösterip göstermediğinin diğer bir ifadeyle DMF gösterme durumlarının belirlenmesi, ilgili testten alınan puanlara göre alınan kararların adil ve yerinde olması bakımından oldukça önemli görünmektedir. Avcu, Tunç ve Uluman (2018) lisans düzeyinde kullanılan bir testin kitapçık türüne göre DMF içerip içermediğini incelemişlerdir. Araştırma sonucunda bazı maddelerde DMF bulunmuştur. Bu durum testlerde maddelerin sıralanışlarının DMF oluşturabildiğini göstermektedir. Ryan ve Chiu (2001) bir matematik testi için maddelerin rastgele ve kolaydan zora sıralandığı iki formu üniversite birinci sınıf öğrencilerine uygulamışlardır. Buna göre maddelerin sıra etkisine göre DMF durumları karşılaştırılmıştır. İki form için maddelerin DMF gösterme durumları bakımından bir fark olmadığı belirlenmiştir. Balta ve Ömür-Sünbül (2017) maddelerin kolaydan zora ve zordan kolay doğru sıralanması ile elde edilen iki matematik test formu için maddelerin sıra etkisine göre DMF durumlarının farklılaşıp farklılaşmadığı incelenmiştir. LR ve MH yöntemleri ile yapılan DMF belirleme süreci sonucunda maddelerin sıra etkisine göre DMF gösterdiği belirlenmiştir. Chiu ve Irwin (2011), tarih, coğrafya ve ekonomiye ilişkin 30 maddenin yer aldığı testte maddeleri derste işleniş sıralarına, kronolojik olarak geçmişten günümüze ve günümüzden geçmişe sıraladıkları üç form için maddelerin DMF gösterip göstermediğini MH ve Breslow-Day yöntemleri ile incelemişlerdir. Buna göre formlar arasında DMF gösteren maddelerin bulunduğu sonucuna ulaşmışlardır. Chiu (2012), bir matematik testinde maddelerin sadece rastgele yerlerini değiştirerek hazırladığı formlarını alan gruplara göre MH ve Breslow-Day yöntemleri ile DMF analizleri yürütmüş ve DMF gösteren maddelerin bulunduğu sonucuna ulaşmıştır.

Maddelerin farklı sıralandıkları formların kullanıldığı sınavlardan biri de Türkiye'de yapılan ortaöğretime veya yükseköğretime geçiş sınavlarıdır. Bu sınavlar, eğitimde bir üst kademeye geçme, bir işe yerleştirme için seçme amacıyla yapılan, testi alan birey için kritik sınavlardır. Bu sınavlardan alınan puanlara göre bireyler

hakkında oldukça önemli kararlar verilmektedir. Milli Eğitim Bakanlığı (MEB) 2018 yılından başlayarak; fen liseleri, sosyal bilimler liseleri, proje uygulayan eğitim kurumları ile mesleki ve teknik Anadolu liselerinin Anadolu teknik programlarına öğrenci yerleştirilmesi amacıyla bir merkezi sınav yapmaktadır. Bu sınavla birlikte çok sayıda öğrencinin nitelikli bir ortaöğretim kurumuna geçiş için karar verilmektedir. Ortaöğretime geçiş sınavında yer alan maddelerin, maddelerin sıralanışı bakımından farklı formlarını alan gruplara göre DMF gösterme durumlarının incelenmesi, ilgili teste göre alınan kararların yerinde olabilmesi açısından gerekli görülmüştür. Buna göre bu araştırmanın temel amacı DMF analizlerinde difR paket programının kullanımına ilişkin izlenen adımları, MH, LR, MTK-OO ve SIBTEST teknikleri kullanılarak elde edilen analiz çıktıları ve yorumu ile bir bütün olarak ele almak bu doğrultuda kullanıcılara yol göstermektir. Bununla birlikte bu temel amaç doğrultusunda Ortaöğretime Geçiş merkezi sınavında yer alan fen maddelerinin de madde sıra etkisine göre DMF durumlarının incelenmesi amaçlanmaktadır.

Yöntem

Araştırmanın bu bölümde araştırma modeli, çalışma grubu, verilerin elde edilmesi ve analizi başlıklarına yer verilmiştir. Ayrıca, bu araştırma 01.01.2020 tarihinden önce yapıldığı için etik kurul kararı zorunluluğu taşımamaktadır.

Araştırmanın Modeli

Bu çalışmada var olan bir durumun var olduğu şekliyle ortaya konulması amaçlandığından tarama modelinde bir çalışmadır (Fraenkel ve Wallen, 2006).

Çalışma Grubu

Milli Eğitim Bakanlığı tarafından, 8.sınıf öğrencileri için, Sınavla Öğrenci Alacak Ortaöğretim Kurumlarına İlişkin Merkezî Sınavı gerçekleştirilmektedir. 2018 yılında gerçekleştirilen sınavda dört tür kitapçık bulunmaktadır. Bu kitapçıklardan rastgele seçilen ikisini alan toplam 2383 öğrenci çalışma grubunu oluşturmaktadır. Araştırma için seçilen kitapçıklar X ve Y olarak yeniden adlandırılmıştır. Buna göre her iki kitapçığı yanıtlayan öğrencilerin aldıkları kitapçık türü ve cinsiyete göre dağılımları Tablo2’de yer almaktadır.

Tablo 2

Çalışma Grubunda Yer Alan Öğrencilerin Kitapçık Türü ve Cinsiyete Göre Dağılımı

		Cinsiyet			
		Kız	Erkek	Toplam	
Kitapçık Türü	X	n (%)	551 (45.6)	657 (54.4)	1,208 (100.0)
		% Toplam	23.1	27.6	50.7
	Y	n (%)	579 (49.3)	596 (50.7)	1,175 (100.0)
		% Toplam	24.3	25.0	49.3
Toplam		n (%)	1,130 (47.4)	1,253 (52.6)	2,383 (100.0)

Tablo 2 incelendiğinde X kitapçığını alan öğrenci sayısının 1208, Y kitapçığını alan öğrenci sayısının 1175 olduğu görülmektedir. Buna göre X ve Y kitapçığını alan öğrenci yüzdeleri %50.7 ve %49.3'tür. DMF araştırılan gruplardaki örneklem büyüklüğünün birbirine yakın değerler olması DMF analizleri bakımından önemli görülmektedir (Herrera ve Gomez, 2008). Çalışma grubu cinsiyet bakımından incelendiğinde kız öğrencilerin grubun %47.4'ünü (1130 öğrenci), erkek öğrencilerin ise grubun %52.6'sını (1253 öğrenci) oluşturduğu görülmektedir.

Verilerin Elde Edilmesi

Araştırma verileri, Milli Eğitim Bakanlığı Ölçme Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü'nden araştırma için gerekli izin süreçleri yürütülerek elde edilmiştir. Merkezi sınav maddeleri incelendiğinde maddelerin dört seçenekli, çoktan seçmeli ve ikili (1-0) puanlanan maddeler olduğu görülmektedir. Doksan maddeden oluşan merkezi sınav sözel ve sayısal olmak üzere iki bölümden oluşmaktadır. Sözel bölüm, 8'inci sınıf Türkçe, din kültürü ve ahlak bilgisi, T.C. inkılap tarihi ve Atatürkçülük ile yabancı dil, sayısal bölüm ise matematik ve fen bilimleri alt testlerinden oluşmaktadır. Araştırma kapsamında 20 maddeden oluşan fen alt testi ele alınmıştır. Merkezi sınav uygulamasında toplam dört kitapçık türü yer almaktadır. Maddeler her yıl yeniden hazırlanmaktadır ve yayımlanmaktadır. Maddelerin kitapçıklara yerleştirilmesinde kullanılan yöntemler bilinmemekle birlikte madde istatistikleri önceden bilinmediğinden bu indekslere göre sıralanamamaktadır. Maddelerin farklı sıralanışına göre oluşturulan kitapçıklar öğrencilere rastgele dağıtılmaktadır. Bu çalışmada örnek DMF analizi çalışması maddelerin farklı sıralanışına dayalı oluşturulan kitapçık türlerine göre DMF gösterme durumlarının incelemesi olduğundan Ortaöğretime Geçiş Sınavı kapsamında yer alan iki kitapçık rastgele seçilmiştir.

Verilerin Analizi

Araştırma kapsamında MTK ve KTK'ya dayalı DMF belirleme teknikleri bir arada kullanıldığından her iki kuramın DMF yöntemlerinde gerektirdikleri varsayımlar incelenmiştir. Normal dağılım, tek boyutlu ve çok boyutlu uç değerlerin ve kayıp verilerin incelenmesi tüm çok değişkenli istatistiklerde yer alan varsayımlardır. Verinin normal dağılımının incelenmesinde X ve Y kitapçığını alan bireylerin testten aldıkları toplam puanların basıklık ve çarpıklık değerleri incelenmiştir. X ve Y kitapçığını alan gruplardan elde edilen puanların sırasıyla çarpıklık katsayıları 0.29 ve 0.23; basıklık katsayıları -0.67, -0.59 olarak hesaplanmıştır. Buna göre hesaplanan çarpıklık, basıklık katsayılarının ± 1 aralığında olması dolayısıyla normalden aşırı sapma göstermeyen bir dağılım olduğu söylenebilir (Büyüköztürk, 2004).

Tek boyutlu uç değerlerin incelenmesi için her bir madde için z standart puanları hesaplanmıştır. Buna göre testi alan bireylerin z puanları -2.37 ile 2.59 arasında değer almaktadır. Hesaplanan z puanlarının ± 3 aralığında yer alması tek değişkenli uçdeğerin bulunmadığını göstermektedir. Çok boyutlu uç değerlerin incelenmesi için,

her bir denek için testteki maddeler üzerinden Mahalanobis uzaklıkları için $\alpha=.001$ ve K-1 serbestlik derecesi için kritik ki-kare değeri ile karşılaştırılmıştır (Mertler ve Vannatta, 2005; Tabachnick ve Fidell, 2007). Hesaplanan Mahalanobis uzaklıkları incelendiğinde 37.53 ile 8.62 arasında değer aldıkları görülmektedir. Testte 20 madde olduğundan serbestlik derecesi 19 için kritik ki-kare değeri 43.82 olarak belirlenmiştir. Buna göre kritik ki-kare değerini aşan Mahalanobis uzaklığı olmadığından dağılımda çok değişkenli uçdeğere rastlanmadığı belirtilebilir.

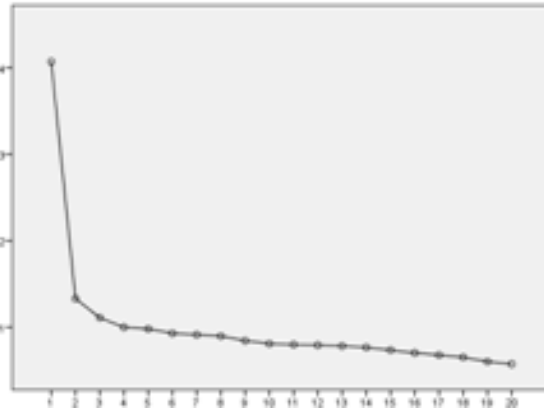
Van Buuren (2012) kayıp veri oranı %30'dan düşük ise atama yapılabileceğini, %30 ve üzerinde ise kayıpları silmek gerektiğini belirtmektedir. Buna göre kayıp veri oranı %30 ve üzerinde olan birey ya da madde olmadığından, analizlerin kayıp verilerden olumsuz yönde etkilenmemesi için boş bırakılan madde yanıtları 0 olarak yeniden kodlanmıştır.

Madde tepki kuramına dayalı DMF yöntemleri bakımından önemli varsayımlardan biri olan tek boyutluluk varsayımı için araştırma verileri üzerinde Açıklayıcı Faktör Analizi (AFA) yapılmıştır. Bu varsayım için faktör analizi sonucu elde edilen faktörlere ilişkin özdeğer ve yamaç-birikinti grafiğinin incelenmesi önerilmektedir (Cattell, 1966; Loehlin, 1987). AFA sonuçlarına göre elde edilen faktör özdeğerleri Tablo 3'te, yamaç birikinti grafiği ise Şekil 1'de yer almaktadır.

Tablo 3

Araştırmada İncelenen Teste İlişkin Faktör Analizi Sonuçları

Faktör Sayısı	Özdeğerler	Açıklanan Varyans	Açıklanan Toplam Varyans
1	4.069	20.347	20.347
2	1.332	6.659	27.006
3	1.112	5.558	32.564
4	1.001	5.003	37.567



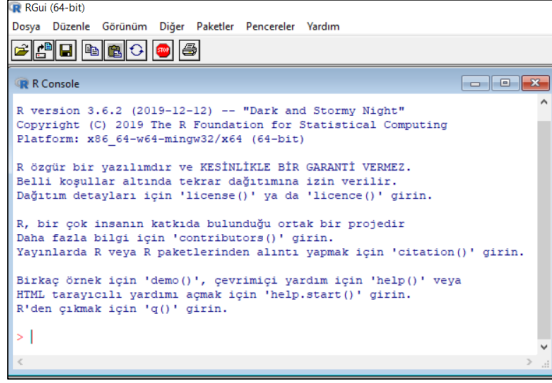
Şekil 1. Yamaç birikinti grafiği

Tablo 1 incelendiğinde AFA sonuçlarına göre birinci faktöre ait özdeğer ile ikinci faktör özdeğeri arasındaki farkın büyük ve ikinci faktör ile diğer faktörler arasındaki farktan çok daha büyük olduğu gözlenmektedir. Bu durum tek boyutluluk varsayımının karşılandığını göstermektedir (Gierl, 2000; Hambleton ve Swaminathan, 1989). Şekil 1 incelendiğinde ise birinci faktörden sonra ani bir düşüş olduğu, ikinci ve diğer faktörlerden sonra ise eğrinin plato yaptığı görülmektedir. Diğer bir ifadeyle birinci faktörden sonra diğer faktörlerin varyansa katkılarının birbirine yakın ve birinci faktöre göre oldukça düşük olduğu görülmektedir.

MTK'ya ait bir diğer varsayım ise yerel bağımsızlık varsayımdır. Yerel bağımsızlık, her bir maddeye verilen yanıtın, diğer maddelere verilen yanıtlardan bağımsız olarak verilmesi olarak tanımlanmaktadır (Crocker ve Algina, 1986). Diğer bir ifadeyle bireylerin bir maddeye ait performansının, diğer maddelerde gösterilen performanslardan olumlu veya olumsuz yönde etkilenmemesi durumuna yerel bağımsızlık denir (Hambleton ve Swaminathan, 1989). Bununla birlikte tek boyutluluk varsayımının karşılandığı durumda aynı yetenek düzeyindeki bireyler için maddelere verilen yanıtlar arasındaki kovaryansın sıfır olduğu bilinmektedir. Buna göre tek boyutluluk varsayımı karşılanmış ise yerel bağımsızlık varsayımı da karşılanır yorumu yapılmaktadır (Hambleton ve Swaminathan, 1989; Hambleton, Swaminathan ve Rogers, 1991). Bu durum, araştırma verileri için tek boyutluluk ve yerel bağımsızlık varsayımlarının karşılandığını göstermektedir.

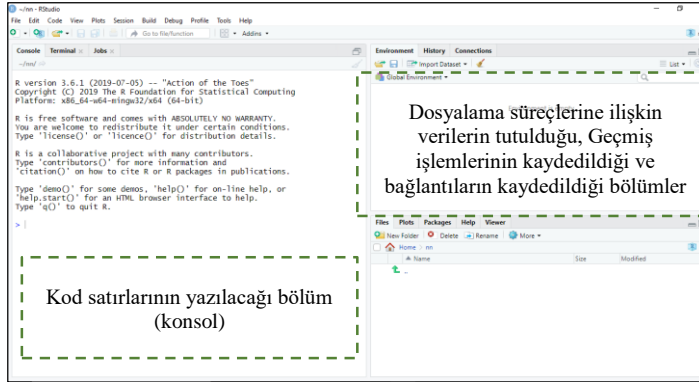
MTK'ya dayalı analizler için veri setinin 1, 2 ve 3 parametrelili lojistik modellerden hangisi ile uyumlu olduğu incelenmelidir. Bu amaçla R yazılımı kullanılarak model veri uyumu analizleri gerçekleştirilmiştir. Bu aşamadan sonra R uygulamaları başlayacağından, öncelikle R yazılımının kurulumuna, veri setinin uygun formata getirilmesine ve model veri uyumu için işlem adımlarına sırasıyla değinilmiştir.

1. R ve R Studio kurulumu. R açık kaynak kodlu ve ücretsiz erişilebilir bir yazılım olması dolayısıyla resmi web sitesinden (<https://cran.r-project.org/>) indirilebilir. Yazılım indirildiğinde Şekil 2'deki gibi bir arayüze sahip "R Konsol" kurulmuş olmaktadır.



Şekil 2. R yazılımı arayüzü

Konsol uygulamaları grafiksel olmayan bir arayüze sahip olduğundan R konsol kullanıcı dostu bir çalışma ortamı sunmaz. Bu nedenle grafiksel öğeler içeren, kullanıcı dostu arayüz sunan R Studio entegre programının da ilgili web sitesinden (RStudio, 2020) indirilerek kurulması önerilmektedir. R Studio, R programlama dili için bir entegre geliştirme ortamıdır. R Studio, R yazılımının çeşitli bileşenlerini (konsol, kaynak düzenleme, grafikler, tarihçe, yardım vb.) kesintisiz ve üretken bir platformda birleştirmeyi amaçlayan açık kaynaklı bir projedir (Allaire, 2012). Şekil 3'te R Studio arayüzü görülmektedir. Bundan sonraki işlemlerin tamamı R Studio'da yapılmıştır.



Şekil 3. R Studio arayüzü

R Studio kurulumu tamamlandıktan sonra henüz R, Windows işletim sisteminde gerekli analizler için hazır değildir. R'de çalışmaya devam etmeden önce Windows'ta çalışabilmek için R araçlarının (Rtools) indirilmesi ve kurulması gerekmektedir (Using Rtools4.0 on Windows, 2020).

2. Paket programların indirilmesi. R çok farklı alanlara ilişkin analizler için ortam sağlamaktadır. Bu nedenle belirli bir konu altında toplanabilecek kodları içeren ve sisteme tanıtılması gereken programlama paketleri vardır. Model veri uyumu incelemeleri için "ltm" (Latent Trait Models under Item Response Theory) paket programının kurulması gerekmektedir. Bu amaçla Şekil 4'te gösterilen kod satırlarının yazılacağı konsola `install.packages("ltm")` kodu yazılarak Enter tuşuyla çalıştırılır. Eğer kodu bir yerden, örneğin bu rapordan kopyalayarak R Studio arayüzüne yapıştırırsanız, hata alabilirsiniz. Yapılan kopyalamalarda özellikle tırnak işareti gibi (") karakterlerin yapıştırılan kod üzerinden düzenlenmesi önerilmektedir. Böylece bilgisayara R kütüphanesine "ltm" ile ilgili tüm programlama dili ve kalıpları eklenmektedir. R kütüphanesinde bulunan bir paketi kütüphaneden çağırmak ve çalışmak için ise `library(ltm)` komutu kullanılmaktadır. Buna göre "ltm" paketini çalıştırmak için konsola `library(ltm)` kodu yazılarak ilgili pakete göre yapılacak işlemler için program hazırlanmış bulunmaktadır.

3. Verinin düzenlenmesi. Analizler için R yazılımında verinin belirli bir formatta olması gerekmektedir. MS Excel'de, her bir maddenin adı (X1, X2, X3...X20) ile birlikte madde yanıtları 1-0 olarak kodlanmalıdır. Hazırlanan Excel dosyası Farklı kaydet seçeneği ile CSV (virgülle ayrılmış) kayıt türünde kaydedilmektedir. Şekil 4'te "veri.csv" adı verilen veri dosyasının bir bölümüne yer verilmiştir.

	A	B	C	D	E	F	G	H	I	J	K
1	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
2	1	0	1	1	1	1	1	1	1	1	1
3	1	1	1	0	0	1	0	0	0	0	0
4	1	0	1	1	0	0	0	1	1	0	0
5	0	0	1	0	0	0	0	0	0	0	1
6	1	0	1	0	0	0	0	0	0	1	0
7	1	0	1	0	0	1	0	0	0	0	0
8	1	1	1	1	0	0	1	1	1	1	1
9	1	1	1	1	1	1	1	1	1	1	1

Şekil 4. Analize hazırlanmış verinin(veri.csv) görünümü

4. Çalışma dizininin belirlenmesi. Analiz sonuçlarının kaydedilmesi ve veri setinin bulunduğu konumun belirli olması amacıyla bilgisayarın istenen bir dizininde bir klasör dosyası oluşturularak veri dosyasının içine kaydedilmesi gerekmektedir. Bu çalışmada bilgisayarın D sürücüsünde DMF adlı bir klasör oluşturularak veri.csv adlı veri dosyası da klasöre kopyalanmıştır. Analizlerin hangi konumda çalıştırılarak kaydedileceğinin R yazılımında da belirlenmesi istenir. Bu amaçla konsol bölümüne `setwd("D:/DMF/")` kodu yazılarak bu işlem gerçekleştirilmiş olmaktadır.

5. Verinin aktarılması. Analizler öncesinde veriler, DMF klasöründe yer alan *veri.csv* adlı 20 maddelik veri dosyasından çekilerek R yazılımında kullanılmak üzere hazır hale getirilir. Bunun için

`m1=read.csv('veri.csv', header=1, sep=";") [,1:20]` kodu yazılarak çalıştırılır.

`header=1` : ilk satırının başlık olduğunu

`sep=";"`: verilerin virgülle ayrıldığını

`[,1:20]`: veri.csv dosyasında yer alan 1. sütundan 20. sütuna kadar olan veriler için *m1* isimli veri matrisinin oluşturulduğu anlamını taşımaktadır. R yazılımında analizler oluşturulan bu veri matrisi üzerinden yürütülmektedir. R Studio'da, sağ penceredeki *Environment* penceresine tıkladığınızda *m1* matrisinin oluşturulduğunu görebilirsiniz. Matristeki verileri de bu pencerenin Data alanındaki *m1*'e tıkladığınızda ya da kod satırına *View (m1)* yazdığınızda *Source* penceresinden görebilirsiniz (Şekil 5).

	X1	X2	X3	X4	X5	X6
1	1	0	1	1	1	1
3	1	0	1	1	0	0
7	1	1	1	1	0	0
8	1	1	1	1	1	1
9	1	1	1	1	1	1
12	1	1	1	1	0	1

Şekil 5. Oluşan matrisin görüntülenmesi

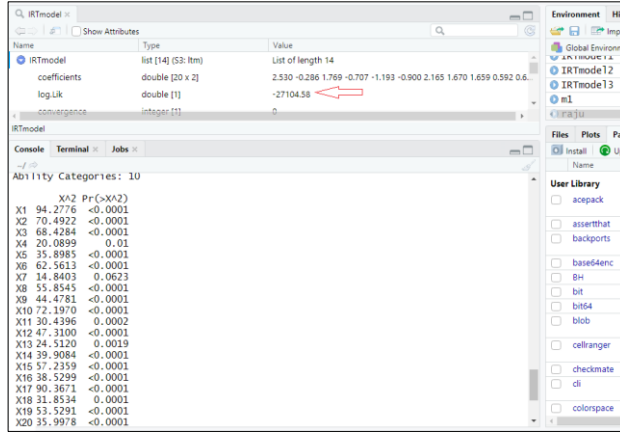
6. Model veri uyumu incelemesi. Gerekli paket programları yükledikten sonra ve veri matrisi oluşturulduktan sonra DMF analizlerine başlanabilir. Veri setinin 1PL, 2PL ve 3PL modeller için uyumunun incelenmesi için 1PL, 2PL ve 3PL modellere ilişkin aşağıda yer alan kodlar sırasıyla çalıştırılmalıdır:

```
IRTmodel1 = rasch(m1, IRT.param = TRUE)
IRTmodel2 = ltm(m1~ z1, IRT.param = TRUE)
IRTmodel3 = tpm(m1, type = "latent.trait", IRT.param = TRUE)
```

Kodların çalıştırılması ile her bir model için log olabirlik (log-likelihood [LL]) değerleri elde edilmektedir. Bununla birlikte her bir model için madde model uyumlarını incelemek üzere aşağıda yer alan kodlar kullanılmaktadır:

```
item.fit(IRTmodel1)
item.fit(IRTmodel2)
item.fit(IRTmodel3)
```

Şekil 6'da item.fit (IRTmodel2) kodunun çalıştırılması ile elde edilen madde model uyum değerleri ve ilgili modele ilişkin değerler yer almaktadır.



Şekil 6. 2PL model için LL değeri ve model madde uyum sonuçları

Her bir model verisine tıklayarak modelin sonuçları görülebilmektedir (Şekil 6). Her bir model için elde edilen LL değerleri ve uyumlu madde sayıları ise Tablo 4'te yer almaktadır.

Tablo 4

LL Değerleri ve Modelle Uyumlu Madde Sayıları

	1PL	2PL	3PL
LL	27675.35	27104.58	26866.11
Modelle Uyum Gösteren Madde Sayısı	19	19	15

Model veri uyumunun değerlendirilmesinde modellere ait LL değerleri arasındaki fark dikkate alınır. Buna göre,

$$LL_{2PL} - LL_{3PL} = 27104.58 - 26866.11 = 238.47$$

$$LL_{1PL} - LL_{2PL} = -27675.35 - 27104.58 = 570.77$$

LL değerleri ki-kare dağılımı göstermektedir. Model veri uyumu için testteki madde sayısına göre kritik ki-kare değeri belirlenmelidir. Buna göre testte 20 madde yer aldığına göre kritik ki-kare değeri $\chi^2_{(20, .05)} = 31.41$ 'dir. Kritik ki-kare değeri ile model uyumlarına ait LL değerleri arasındaki farklar karşılaştırılır. Buna göre 238.47 ve 570.77 fark değerleri kritik ki-kare değerinden (31.41) büyük olduğuna göre testin 3 PL modelle uyumlu olduğu söylenebilir. Ancak model ile uyumlu madde sayıları incelendiğinde 2PL modelle uyumlu madde sayısı daha fazla olduğundan testin 2PL

modelle uyumlu olduğu kararı verilmiştir. Şekil 8'de de görüldüğü üzere 2PL modele ilişkin model uyum istatistiklerine göre 7 no'lu maddenin model ile uyumlu olmaması nedeniyle madde analiz dışı bırakılmıştır. Buna göre veri seti KTK ve MTK varsayımları bakımından incelenmiş ve varsayımların karşılandığı belirlenmiştir.

Değişen madde fonksiyonunun belirlenmesi. Bu aşamadan sonra testte yer alan maddelerin DMF gösterme durumlarını incelemek üzere izlenen DMF analiz adımları sunulmuştur. Veri setinden bir maddenin çıkarılmasıyla kalan 19 madde için kitapçık türüne göre DMF analizleri yürütülmelidir. Bu noktada, ilgili maddeye ilişkin verilerin yer aldığı sütun veri.csv dosyasından çıkarılmalıdır. Ayrıca kitapçık X'i alanların 1, kitapçık Y'yi alanların 0 olarak kodlandığı grup değişkenine (G) ilişkin verilerinde eklenmesiyle veri.csv dosyası güncellenmiştir (Şekil 7).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	X1	X2	X3	X4	X5	X6	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	G
2	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	1	1	1
3	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
4	1	0	1	1	0	0	0	1	1	0	0	0	1	1	1	1	0	1	1	1
5	0	0	1	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0	1	1
6	1	0	1	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	1	1
7	1	0	1	0	0	1	0	0	0	0	0	0	1	1	0	1	0	0	0	1

Şekil 7. Grup değişkeni eklenmiş verinin (veri.csv) görünümü.

Şekil 7'de yer alan veri dosyasının R yazılımında kullanılabilir olabilmesi için daha önce de belirtildiği gibi veriler R yazılımına aktarılmalıdır. Bu amaçla aşağıda verilen kod kullanılmıştır. Kodda madde sayısının 19'a düştüğüne dikkat edilmelidir. Böylece DMF analizleri için *m2* adlı matris dosyası oluşturulmaktadır.

```
m2=read.csv('veri.csv', header=1, sep=";") [1:19]
```

Bir sonraki adımda veri dosyasındaki grup değişkeninin veri dosyasının hangi sütununda yer aldığı belirtilmelidir. Kod ile veri.csv dosyasının G başlıklı sütununun "g" matrisi olarak R yazılımında kullanılabilir olması sağlanmaktadır.

```
g=read.table('veri.csv', header=1, sep=";") [,c('G')]
```

Verilerin R'a aktarılması ile veri seti DMF analizler için hazır duruma gelmiştir. Analizlerin hangi konumda çalıştırılarak kaydedileceğinin R yazılımında da belirlenmesi istenir. Bu amaçla konsol bölümüne `setwd("D:/DMF/")` kodu yazılarak bu işlem gerçekleştirilmiş olmaktadır. DMF analizleri için 2.adımda bazı paket programların indirilmesi gerekmektedir. *diffR* paketi DMF analizleri için gerekli olan temel paket programdır. Bununla birlikte *ltm* paketi MTK'ya dayalı analizler için *lme4* paketi, karma (doğrusal, genelleştirilmiş doğrusal ve doğrusal olmayan) modellerin analizi ve uyumu için *mirt* (Multidimensional Item Response Theory)

paketi ise yine çok boyutlu MTK analizleri için yüklenmektedir. Buna göre ilgili paketlerin yüklenmesine ilişkin kodlar aşağıdaki gibidir:

```
install.packages("difR")
install.packages("lrm")
install.packages("lme4")
install.packages("mirt")
```

Belirtilen kodlar ile paketler yüklendikten sonra üzerinde çalışabilmek için ilgili paket programlara ilişkin kütüphanelerin çalıştırılması gerekmektedir. Bu amaçla aşağıdaki kodlar kullanılmalıdır:

```
library(difR)
library(lrm)
library(lme4)
library(mirt)
```

DMF analizleri için gerekli paketler indirildikten sonra ilgili DMF belirleme yöntemine ilişkin kodlar ile analizler gerçekleştirilebilmektedir. Bu araştırmada LR, MH, MTK-OO ve SIBTEST teknikleri ile DMF analizleri yürütülmüştür. difR'da analiz edilebilen tüm yöntemler için aynı parametreler kullanılabilirdiği gibi (*data*, *group*, *focal.name*, *anchor*, *output* vb.) farklı parametreler de söz konusu olabilmektedir (Magis, Beland ve Raiche, 2018). Buna göre her bir DMF yöntemine ilişkin analiz için kullanılan kodlar Tablo 5'te yer almaktadır.

Tablo 5

Farklı Yöntemlere Göre DMF Analizleri İçin Yazılması Gereken Kod Satırları

DMF Yöntemi	Kod Satırı
MH	difMH(m2, group=g, focal.name = 1, MHstat = "MHChisq", correct = TRUE, exact = FALSE, alpha = 0.05, save.output = 1, output = c("difMH", "default"))
LR	difLogistic(m2, group=g, focal.name=1, match = "score", type = "both", criterion = "LRT", alpha = 0.05, save.output = 1, output = c("difLojistik", "default"))
OO	difLRT(m2, group = g, focal.name = 1, save.output = 1, output = c("difOO", "default"))
SIBTEST	sibTest(m2,g)
Farklı Teknikleri Karşılaştırma	dichoDif(m2,group=g,focal.name=1, method=c("LRT","MH","Logistic","SIB"), save.output=1, output= c("karsilastirma", "default"))

Tablo 5'te yer alan kod satırları için genel olarak aşağıdaki açıklamalar yapılabilir (Magis, Beland ve Raiche, 2018):

- Her bir yöntem için öncelikle maddelerin yer aldığı veri matrisinin adı *m2*, daha sonra grup değişkeni için oluşturulan veri matrisi *group = g* kodunun tüm yöntemlerde yer aldığı görülmektedir.

- Grup değişkeninde hangi grubun odak grup olduğu ise *focal.name* = 1 kodu ile belirtilmektedir. Buna göre veri setinde “1” ile kodlanan grubun odak grup olduğu belirtilmiş olmaktadır.
- *MHstat* DMF analizi için kullanılacak istatistiği belirtir. Olası değerler *MHChisq* (*Mantel-Hanzel*) (*varsayılan*) ve *logOR* (*Olabilirlik Oran*)’dır.
- *Correct=TRUE* Süreklilik düzeltmesi kullanılmalı mı? (*varsayılan*: EVET)
- *exact=FALSE* Kesin bir test hesaplanmalı mı? (*varsayılan*: HAYIR).
- *match=score* ölçüt türünü belirtir. Test puanı ise *score* olur. Farklı sürekli bir değişken de olabilir. Burada test puanı alınmıştır.
- *type = both*, kodu için varsayılan *both* değeridir, “*udif*” tek biçimli ve “*nudif*” tek biçimli olmayan DMF gösteren maddelerin belirlenebilmesi için kullanılmaktadır.
- *criterion = LRT* kodu hangi DMF istatistiğinin kullanılacağını belirtir. Varsayılan *LRT* iken diğer bir seçenek ise *Wald* istatistiğidir.
- Analiz sonuçlarının kaydedilmesi komutu ise *save.output = 1* kodu ile sağlanmaktadır. Bu koddan sonra hangi isimle kaydedilmek isteniyorsa belirtilmelidir. Örneğin MH yöntemi için yazılan “*output = c(“difMH”, “default”)*” kodu ile analiz sonuçlarının “*difMH*” adıyla kaydedilmesi sağlanmaktadır.
- Bununla birlikte anlamlılık düzeyinin *alpha = 0.05* koduyla belirtildiği görülmektedir.
- Madde Tepki Kuramı’na dayalı modellerde ise veri setinin uyumlu olduğu modelin belirtilmesi gerekmektedir. Bunun için ise *model = “2PL”* kodu kullanılmıştır.
- SIBTEST yönteminde veri matrisi ve grup bilgisini girilmesi yeterlidir.
- Farklı tekniklerin karşılaştırılması için *dichoDif* ana kodu içerisinde farklı olarak hangi yöntemler için analizlerin karşılaştırılacağını belirtilmesi gerekmektedir.
- Tablo 5’te verilen her bir kod satırı çalıştırılarak her bir yöntem için DMF sonuçları incelenebilir. Kod satırları çalıştırılmasıyla analiz tamamlanmıştır. Analiz sonuçları *output* parametresinde belirtilen dosya adında, daha önce belirtilen konuma (D:/DMF) kaydedilmiştir.

Bulgular

Bu başlık altında araştırmada kullanılan her bir DMF belirleme tekniği için elde edilen analiz çıktılarına doğrudan yer verilmiştir. Analiz çıktıları ile birlikte elde edilen bulguların yorumları da bu başlık altında ele alınmıştır.

Mantel-Haenszel Yöntemine göre DMF Analizi Sonuçları

MH yöntemi DMF analiz sonuçları Şekil 8’de sunulmuştur. Çıktı dosyasında parametrelerle ilgili açıklamalardan sonra MH ki-kare istatistikleri (*Stat.*) ve anlamlılık düzeyleri (*p-value*) sunulmuştur.

```

Detection of Differential Item Functioning using Mantel-Haenszel
method with continuity correction and without item purification
Results based on asymptotic inference
Matching variable: test score
No set of anchor items was provided
No p-value adjustment for multiple comparisons
Mantel-Haenszel chi-square statistic:
Stat.  P-value
X1  1.8502 0.1738
X2  0.3022 0.5825
X3  0.3439 0.5576
X4  0.1887 0.6640
X5  0.0078 0.9298
X6  0.3789 0.5382
X8  2.9642 0.0851
X9  0.6518 0.4195
X10 0.0697 0.7918
X11 0.4706 0.4927
X12 0.5205 0.4706
X13 0.3896 0.5325
X14 1.4024 0.2363
X15 0.0001 0.9910
X16 2.4831 0.1151
X17 0.4593 0.4980
X18 0.9772 0.3229
X19 2.1644 0.1412
X20 0.8190 0.3655
Signif.codes: 0 '****' 0.001 '***' 0.01 '**' 0.05 '.' 0.1 ' ' 1
Detection threshold: 3.8415 (significance level: 0.05)
Items detected as DIF items: No DIF item detected

```

Şekil 8. MH yöntemi DMF analizi çıktısı

Şekil 8'e göre maddelere ait ki kare testi sonuçları hiçbir düzeyde anlamlı değildir ($p > .05$). Başka bir ifadeyle hiçbir madde DMF göstermemektedir. Bunun bir sonucu olarak da son satırda hiçbir maddede DMF görülmediği notu düşülmüştür. MH yöntemi DMF analizi sonuçlarının DMF büyüklükleriyle ilgili değerler de yine Şekil 9'da yer almaktadır. Şekil 9'da verilen analiz sonuçları incelendiğinde MH odds oranı kestirimi α_{MH} sütununda, etki büyüklüğü ise Δ_{MH} (Δ_{MH}) sütununda verilmektedir (Magis ve diğ., 2010). Elde edilen Δ_{MH} istatistiği DMF derecesini ortaya koyan ve DMF etki büyüklüğü olarak adlandırılan bir değerdir. Maddelere ait Δ_{MH} değerinin negatif olması referans grup için, pozitif olması ise odak grup için bir avantaj sağladığı anlamını taşımaktadır. Örnek araştırmada odak grup X kitapçığını alanlar olarak belirlenmiştir. Bununla birlikte örneğin madde1 için Δ_{MH} değerinin pozitif olması odak grup için avantaj sağladığı biçiminde yorumlanabilecektir.

Alanyazında DMF düzeylerinin farklı sınıflandırılmalarının olduğu görülmektedir. DMF düzeyi kategorilerinin ETS Delta Ölçeği'ne göre yapıldığı belirtilmektedir. ETS Delta Ölçeği ETS'nin (Educational Test Service) belirlediği kategori aralıklarıdır. Şekil 9'da MH yöntemi DMF analizi çıktısının devamı sunulmuştur.

```

Effect size (ETS Delta scale):
Effect size code:
'A': negligible effect
'B': moderate effect
'C': large effect
alphaMH deltaMH
X1 0.8328 0.4301 A
X2 1.0646 -0.1471 A
X3 0.9285 0.1743 A
X4 0.9566 0.1043 A
X5 1.0145 -0.0338 A
X6 1.0712 -0.1617 A
X8 0.8376 0.4165 A
X9 0.9246 0.1842 A
X10 1.0343 -0.0792 A
X11 0.9301 0.1703 A
X12 1.0774 -0.1753 A
X13 1.0756 -0.1713 A
X14 1.1326 -0.2926 A
X15 1.0038 -0.0089 A
X16 0.8538 0.3715 A
X17 0.9235 0.1870 A
X18 1.1087 -0.2425 A
X19 1.1566 -0.3418 A
X20 1.0963 -0.2161 A
Effect size codes: 0 'A' 1.0 'B' 1.5 'C'
(for absolute values of 'deltaMH')
output was captured and saved into
file'C:/Users/TOSHIBA/Desktop/dif/difMH.txt'

```

Şekil 9.MH yöntemi DMF analizi çıktısının devamı

Şekil 9 incelendiğinde bu kategorilerin aralıklarının *Effect size codes: 0 'A' 1.0 'B' 1.5 'C'* olarak sunulduğu görülmektedir. Sonuç olarak MH yöntemine göre tüm maddelerde, DMF yoktur veya ihmal edilebilir düzeydedir (A düzeyi). Analiz sonuçları, grupların her birinin lehine çalışan ve DMF düzeyi temelinde sınıflandırılarak ve madde numaraları tablolaştırılarak sunulabilir. Bu örnekte hiçbir madde DMF göstermediğinden böyle bir tablo sunmak olanaklı değildir. Daha ayrıntılı değerler sunmak istenen bir araştırmada, her bir madde için odds oranları, Δ MH değerleri, DMF düzeyleri de sunulabilir.

Lojistik Regresyon Yöntemine Göre DMF Analizi Sonuçları

LR yöntemi DMF analizi sonuçları Şekil 10'da sunulmuştur. Şekilde analiz parametre değerleri ile ilgili açıklamalardan sonra LR istatistiği (*Stat.*) ve anlamlılık düzeyleri (*p-value*) sunulmaktadır.

```

Detection of both types of Differential Item Functioning using Logistic
regression method, without item purification and with LRT DIF statistic
Matching variable: test score No set of anchor items was provided
No p-value adjustment for multiple comparisons

```

Logistic regression DIF statistic:

```

Stat. P-value
x1 4.7268 0.0941
x2 0.3594 0.8355

```

```
x3 1.2163 0.5444
x4 0.4189 0.8110
x5 0.3132 0.8550
x6 1.7369 0.4196
x8 3.9082 0.1417
x9 1.2134 0.5451
x10 3.9959 0.1356
x11 0.7182 0.6983
x12 1.0512 0.5912
x13 2.6384 0.2673
x14 1.8607 0.3944
x15 0.0703 0.9654
x16 5.9490 0.0511
x17 2.1570 0.3401
x18 0.8669 0.6483
x19 1.3796 0.5017
x20 1.1785 0.5548
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Detection threshold: 5.9915 (significance level: 0.05)
Items detected as DIF items: No DIF item detected
Effect size (Nagelkerke's R^2):
Effect size code:
'A': negligibleeffect
'B': moderateeffect
'C': largeeffect

R^2    ZT JG
x1 0.0015 A A
x2 0.0001 A A
x3 0.0004 A A
x4 0.0002 A A
x5 0.0001 A A
x6 0.0006 A A
x8 0.0011 A A
x9 0.0004 A A
x10 0.0010 A A
x11 0.0002 A A
x12 0.0003 A A
x13 0.0011 A A
x14 0.0006 A A
x15 0.0000 A A
x16 0.0018 A A
x17 0.0007 A A
x18 0.0004 A A
x19 0.0005 A A
x20 0.0004 A A
Effect size codes:
Zumbo& Thomas (ZT): 0 'A' 0.13 'B' 0.26 'C' 1
Jodoin&Gierl (JG): 0 'A' 0.035 'B' 0.07 'C' 1
Output was captured and saved into file
'C:/Users/TOSHIBA/Desktop/dif/difLojistik.txt'
```

Şekil 10. LR yöntemi DMF analizi çıktısı

Şekil 10'a göre hiçbir maddenin DMF sonuçları anlamlı değildir ($p > .05$). Tüm maddelerde A düzeyinde DMF olduğu da gözlenmektedir. Bu durum, LR yöntemine göre ilgili maddelerde DMF olmadığı biçiminde yorumlanabilir. Çıktı dosyasında yer alan *Items detected as DIF items: No DIF item detected* ifadesi de benzer şekilde sonucu göstermektedir.

MTK'ya Dayalı Olabilirlik Oran Yöntemine Göre DMF Analizi Sonuçları

Olabilirlik oran yöntemine göre DMF analizi sonuçları Şekil 11'de sunulmuştur. Şekil 11'de yer alan analiz çıktısında Olabilirlik oran istatistiğine (Likelihood Ratio Statistic) ilişkin istatistik değeri *Stat.* sütununda ve anlamlılık düzeyleri ise *p-value* sütununda sunulmaktadır. Şekle göre anlamlılık düzeyi .05'ten küçük olan değerler* ile belirtilmiştir. Buna göre X8, X14, X16 no'lu maddelerin DMF gösterdiği belirlenmiştir. *Stat.* alanındaki değerleri G^2 istatistiğini göstermektedir. DMF gösterdiği belirlenen X8, X14, X16 no'lu maddelere ilişkin G^2 değerleri ise sırasıyla 4.4213, 4.6874, 4.0893 olarak hesaplanmıştır. Buna göre olabilirlik oran istatistik değerlerinin $3.84 < G^2 < 9.4$ aralığında yer alması ihmal edilebilir düzeyde (A) DMF düzeyini gösterdiğinden (Greer, 2004), maddelerin ihmal edilebilir düzeyde DMF gösterdiği söylenebilir.

```

Detection of Differential Item Functioning using Likelihood Ratio Test
without item purification
No p-value adjustment for multiple comparisons
Likelihood Ratio statistic:
  Stat.  P-value
X1      2.4357  0.1186
X2      1.8972  0.1684
X3      2.0866  0.1486
X4      1.6094  0.2046
X5     -1.9752  1.0000
X6      1.3328  0.2483
X8      4.4213  0.0355 *
X9      2.2249  0.1358
X10     2.4047  0.1210
X11     2.1571  0.1419
X12     2.6610  0.1028
X13     3.0526  0.0806 .
X14     4.6874  0.0304 *
X15     1.2803  0.2578
X16     4.0893  0.0432 *
X17     3.4311  0.0640 .
X18     3.5037  0.0612 .
X19     3.6689  0.0554 .
X20     3.7334  0.0533 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Detection threshold: 3.8415 (significance level: 0.05)
Items detected as DIF items:
X8
X14
X16
Output was captured and saved into file
'C:/Users/TOSHIBA/Desktop/dif/LRTresults.txt'

```

Şekil 11. MTK-OO yöntemine göre DMF analizi çıktısı

SIBTEST Yöntemine Göre DMF Analizi Sonuçları

SIBTEST yöntemine göre DMF analizi sonuçları Şekil 12'de sunulmuştur. Şekil 12'de yer alan SIBTEST analiz sonuçları incelendiğinde $\$Beta$, $\$SE$, $\$X2$, $\$df$, $\$p.value$ değerleri ile sırasıyla beta, standart hata (standart error-SE), χ^2 (ki-kare),

serbestlik derecesi (degree of freedom-df) ve p değerleri yer almaktadır. İlgili değerlere ilişkin bulgular ise diğer analiz çıktılarında olduğu gibi alt alta değil yan yana verilmiştir.

```

$Beta
[1] -0.031927247 0.015259146 -0.012306920 -0.005238659 0.004646948
0.012348480 -0.027621048 -0.010609764
[9] 0.008510391 -0.016582750 0.012830228 0.008189025 0.024915081
0.010507575 -0.024794240 -0.008188328
[17] 0.017391579 0.031743745 0.022377613

$SE
[1] 0.01824609 0.01814779 0.01695322 0.01906388 0.01723816 0.01756688
0.01872007 0.01945864 0.01782877 0.01845683
[11] 0.01885550 0.01630781 0.01806284 0.01888870 0.01870575 0.01682498
0.01757585 0.01879592 0.01901237

$x2
[1] 3.06184729 0.70698978 0.52698038 0.07551234 0.07266977 0.49412613
2.17703444 0.29729421 0.22785373 0.80723319
[11] 0.46301228 0.25215823 1.90262284 0.30945748 1.75691827 0.23685427
0.97914159 2.85225947 1.38533521

$df
[1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

$p.value
[1] 0.08015050 0.40044495 0.46787914 0.78347368 0.78748854 0.48209249
0.14008416 0.58558370 0.63311984 0.36893952
[11] 0.49621953 0.61555949 0.16778503 0.57801329 0.18500937 0.62648742
0.32241083 0.09124562 0.23919445

$type
[1] "udif

```

Şekil 12. SIBTEST yöntemine göre DMF analizi çıktısı

Şekil 12'ye göre her bir satırın başında yer alan numaralar ilgili satırda yer alan ilk değer hangi maddeye ait olduğunu belirtmektedir. Buna göre [1] no'lu maddenin β değeri *0.031927247* olarak belirlenmiştir. Bu değer yanındaki değer ikinci maddeye aittir, bu şekilde birinci satırda sırasıyla maddelere ilişkin değerler yer almaktadır. Yine Beta değerleri için ikinci satırda yer alan [9] bu satırdaki ilk değer 9. maddeye ait olduğunu belirtmektedir. Buna göre 9. madde için beta değeri de *0.008510391*'dir. SIBTEST sonuçlarında yer alan p değerlerinin tamamının .05'ten büyük olması nedeniyle maddelerin hiçbirinde DMF olmadığı yorumu yapılabilir. Eğer DMF belirlenen madde bulunsaydı *beta* değerleri incelenerek DMF düzeyi hakkında alanyazında belirtilen kritik aralıklara göre (Gotzmann ve diğ., 2006; Stout ve Roussos, 1995) karar verilebilecekti. Tüm bu yöntemlerden elde edilen sonuçların sunumunda, her bir yöntemin istatistiksel sonuçlarının verildiği tablolar oluşturulabilir. Ancak DMF analizi araştırmalarında bulgularda her bir yöntemde DMF gösteren maddelerin ve DMF düzeylerinin sunumunun yanında, DMF'nin hangi grup lehine olduğunun da belirtilmesi gerekmektedir.

Tartışma, Sonuç ve Öneriler

Teknolojinin gelişmesi ile çok sayıda istatistiksel analiz kolaylıkla yapılabilmektedir. Bu durum bilimsel araştırma sürecini hızlandırmakta ve analiz sonuçlarında araştırmacıdan kaynaklı hataları azaltmaktadır. Ancak bu hız ve güvenilirlik katkısının gerçekleşmesi, ilgili yazılımın araştırmacılar tarafından etkin biçimde kullanılabilmesiyle olanaklıdır. Alanyazında DMF analizi yapılan çok sayıda araştırma (Asil ve Gelbal, 2012; Bakan-Kalaycıoğlu ve Kelecioğlu, 2011; Bolt ve Ysseldyke, 2008; Karakoç-Alatlı ve Çokluk-Bökeoğlu, 2018; Osadebe ve Agbure, 2018; Ünsal-Özberk ve Koç, 2017) olmasına karşın, bu analizlerin işlem basamaklarını bilimsel çerçevede anlatan bir kaynak olmadığı gözlenmektedir. Ayrıca yapılan araştırmalar incelendiğinde DMF analizlerini yürütebilmek adına tek bir çalışmada birçok istatistiksel programın kullanımı gerekmektedir. Bu nedenle DMF analizlerini tek bir istatistiksel programla yürütebilme kolaylığı sağlayan R paketleri ile DMF analizlerine ilişkin işlem adımlarının, analiz çıktılarının, çıktılarda yer alan bulgular ve yorumlarının bir arada yer aldığı bir çalışma ile DMF analizlerinin yürütüleceği araştırmalara yöntem bilimsel olarak katkı sağlanacağı öngörülmektedir. Araştırmanın, yanlılığa ilişkin incelemelerin yapıldığı araştırmalara destek sunması beklenmektedir. Bu da farklı gruplardan bireylerin karşılaştırılabildiği ölçmelerde, karşılaştırmanın daha anlamlı olmasına ve yerinde kararlar alınmasına hizmet edecektir.

Alanyazın incelendiğinde R yazılımının kullanımına ilişkin (Horgan, 2012; Doğan ve Uluman, 2016) veya bazı istatistiksel analizlerin R yazılımı ile yürütülebilmesi için temel analiz adımları ve analiz çıktılarının ele alındığı çalışmaların olduğu görülmektedir (Luo, Arizmendi ve Gates 2019; Stiglic ve diğ., 2019). Bununla birlikte R yazılımında DMF belirlemeye ilişkin yapılan bir çalışmada MH yöntemi ayrıntılı olarak ele alınmıştır. Alanyazın incelendiğinde R yazılımı farklı analizler üzerinden işlem adımları ve analiz çıktılarının yorumu bakımından çalışma konusu olarak yer almaktadır. Ancak DMF çalışmalarında birçok DMF belirleme tekniğinin bir arada kullanılmasının önerildiği, bununla birlikte birçok DMF çalışmasında birden fazla istatistik programının kullanıldığı düşünüldüğünde R yazılımında DMF belirleme tekniklerinin bir arada ele alındığı bu çalışma oldukça önemlidir. Bu araştırma ile MH, LR, MTK-OO ve SIBTEST DMF belirleme tekniklerine ilişkin işlem adımları, kodlarıyla birlikte sırasıyla ele alınmış, analiz çıktıları ise yorumlarıyla birlikte sunulmuştur. Bu haliyle alanda ve Türkiye'de yapılan benzer bir çalışmaya rastlanmamıştır.

Araştırmada örnek olarak belirlenen veride ele alınan amaç ve sonucu da tartışmaya değer bir konudur. Ortaöğretime geçiş merkezi sınavı fen maddelerinin kitapçık türüne göre genel olarak DMF göstermediği ya da ihmal edilebilir düzeyde DMF gösterdiği gözlenmiştir. Araştırma bu yönüyle Ryan ve Chiu'nun (2001) çalışması ile benzerlik göstermektedir. Ancak madde sıra etkisine bağlı olarak farklı formları alan gruplara göre DMF gösteren maddelerin belirlendiği çalışmalar da bulunmaktadır (Avcu ve diğ., 2018; Balta ve Ömür-Sünbül, 2017; Chiu ve Irwin, 2011; Chiu, 2012).

Bu araştırmada R yazılımı ile MH, LR, MTK-OO ve SIBTEST teknikleri ele alınmıştır. Diğer DMF belirleme teknikleri ile de benzer çalışmalar yürütülebilir. Yalnızca DMF belirleme teknikleri için değil diğer istatistiksel teknikler için de benzer araştırmalar yürütülebilir. R yazılımı ile birçok istatistiksel analiz tek bir yazılım ile yürütülebilmektedir. Bu noktada araştırmacıların farkındalığını arttırmak adına eğitimler düzenlenebilir. Bu araştırmada Ortaöğretime Geçiş Sınavı fen alt testi için madde sıra etkisi bakımından kitapçık türlerine göre maddelerin DMF gösterme durumları incelenmiştir. Farklı sınavların farklı testlerine ilişkin benzer çalışmalar yürütülebilir. Bu araştırmada madde sıra etkisi bakımından DMF gösteren madde belirlenmemiştir. Ancak ilgili değişken bakımından DMF gösteren maddelerin belirlendiği çalışmalar mevcuttur. Madde sıra etkisi bakımından DMF gösterdiği belirlenen maddeler için yanlılık çalışmaları yürütülerek bu durumun nedenleri ortaya konulabilir.

Kaynakça

- Adedoyin, O. (2010). An investigation of the effects of teachers' classroom questions on the achievements of students in mathematics: Case study of Botswana community junior secondary schools. *European Journal of Educational Studies*, 2(3), 313-329. Retrieved from <http://connection.ebscohost.com/c/articles/67192996/investigation-effects-teachers-classroom-questions-achievements-students-mathematics-case-study-botswana-community-junior-secondary-schools>
- Akalın, Ş. (2014). *Kamu personeli seçme sınavı genel yetenek testinin madde yanlılığı açısından incelenmesi* (Yayınlanmamış doktora tezi). <https://tez.yok.gov.tr/UlusalTezMerkezi/>'nden erişilmiştir. (Tez No. 381796).
- Allaire, J. J. (2011, August). *RStudio: Integrated development environment for R*. Paper presented at the R User Conference, useR! University of Warwick, Coventry, UK. Retrieved from https://www.r-project.org/conferences/useR-2011/abstract_booklet.pdf
- Asil, M. ve Gelbal, S. (2012). PISA öğrenci anketinin kültürler arası eşdeğerliği. *Eğitim ve Bilim*, 37(166), 236-249 <http://egitimvebilim.ted.org.tr/index.php/EB/article/view/1501/462> adresinden erişilmiştir.
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education [AERA/APA/NCME]. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Avcu, A., Tunç, E. B., and Uluman, M. (2018). How the order of the items in a booklet affects item functioning: Empirical findings from course level data. *European Journal of Education Studies*, 4(3), 227-239. doi:10.5281/zenodo.1199695

- Bakan-Kalaycıoğlu, D., and Kelecioğlu, H. (2011). Item bias analysis of the university entrance examination. *Education and Science*, 36(161), 3-13. Retrieved from <http://egitimvebilim.ted.org.tr/index.php/EB/article/view/143>
- Balta, E., and Ömür-Sünbül, S. (2017). An investigation of ordering test items differently depending on their difficulty level by differential item functioning. *Eurasian Journal of Educational Research*, 72(2017), 23-42. doi: 10.14689/ejer.2017.72.2.
- Barcikovski, R. S., and Olsen, H. (1975). Test item arrangement and adaptation level. *The Journal of Psychology*, 90(1), 87-93. doi: 10.1080/00223980.1975.9923929.
- Beaujean, A. A. (2013). Factor analysis using R. *Practical Assessment: Research & Evaluation*, 18(4), 1-11. <https://doi.org/10.7275/z8wr-4j42>.
- Bolt, D. (2000). A SIBTEST approach to testing DIF hypothesis using experimentally designed test items. *Journal of Educational Measurement*, 37(4), 307-327. Retrieved from <https://www.jstor.org/stable/1435243?seq=1>
- Bolt, S. E., and Ysseldyke, J. (2008). Accommodating students with disabilities in large-scale testing: A comparison of differential item functioning (DIF) identified across disability types. *Journal of Psychoeducational Assessment*, 26(2), 121-138. doi:10.1177/0734282907307703
- Büyüköztürk, Ş. (2004). *Veri analizi el kitabı*. Ankara: PegemA Yayıncılık.
- Camilli, G., and Shepard, L. A. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: Sage.
- Cattell, R. B. (1966). The data box: Its ordering of total resources in terms of possible relational systems. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology* (pp. 67-128). Chicago, IL: Rand-McNally
- Cheung, G. W., and Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-cultural Psychology*, 31(2), 188-213. Retrieved from <https://doi.org/10.1177/0022022100031002003>
- Chiu, P. (2012, April). The effect of item position on state mathematics assessment. Paper presented at the Annual Meeting of the American Educational Research Association, Kansas. Retrieved from https://aai.ku.edu/sites/cete.ku.edu/files/docs/Presentations/2012_04_ChIU_Math_Item_Ordering_AERA_2012.pdf
- Chiu, P. C., and Irwin, P. M. (2011, April). *Chronological item ordering: Does it make a difference on a state history and government assessment?* Paper presented at the Annual Meeting & Training Sessions of the National Council on Measurement in Education, New Orleans, Louisiana, USA. Retrieved from

<https://docplayer.net/131592073-Chronological-item-ordering-does-it-make-a-difference-on-a-state-history-and-government-assessment-pui-chiu-and-patrick-m.html>

- Crocker, L. M., and Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando, FL: Holt, Rinehart and Winston.
- Çepni, Z. (2011). *Değişen madde fonksiyonlarının SIBTEST, Mantel-Haenzsel, lojistik regresyon ve madde tepki kuramı yöntemleriyle incelenmesi* (Yayınlanmamış doktora tezi). <https://tez.yok.gov.tr/UlusalTezMerkezi/>'nden erişilmiştir (Tez No. 308512).
- Çokluk, Ö., Gül, E., and Doğan-Gül, Ç. (2016). Examining differential item functions of different item ordered test forms according to item difficulty levels. *Educational Sciences: Theory & Practice*, 16, 319-330. <http://dx.doi.org/10.12738/estp.2016.1.0329>
- DeMars, C. E. (2010). Type I error inflation for detecting DIF in the presence of impact. *Educational and Psychological Measurement*, 70(6), 961-972. doi: 10.1177/0013164410366691
- Doğan, C. D. ve Uluman, M. (2016). İstatistiksel veri analizinde R yazılımı ve kullanımı, *İlköğretim Online*, 15(2), 615-634, 2016. doi: 10.17051/io.2016.24991.
- Doğan Gül, Ç. ve Çokluk Bökeoğlu, Ö. (2018). Madde güçlüklerine göre farklı sıralanan testlerde düşük ve yüksek kaygılı öğrencilerin akademik başarılarının karşılaştırılması. *İnönü Üniversitesi Eğitim Fakültesi Dergisi*, 19 (3) , 252-265. doi: 10.17679/inuefd.341477.
- Dorans, N. J., and Holland, P. W. (1993). DIF detection and description: Mantel-Haenzsel and standardization. In P. W. Holland and H. Wainer (Eds.), *Differential item functioning* (pp. 35–66). New Jersey: Lawrence Erlbaum Associates Publishers.
- Er, F. ve Sönmez, H. (2005). Temel istatistik eğitiminde R for Windows paket programı kullanımı. *Eskişehir Osmangazi Üniversitesi Mühendislik ve Mimarlık Fakültesi Dergisi*, 18(2), 71-86. <https://dergipark.org.tr/tr/download/article-file/321059> adresinden erişilmiştir.
- Fraenkel, J. R., and Wallen, N. E. (2006). *How to design and evaluate research in education*. New York, NY: McGraw Hill Higher Education.
- Gierl, M. J., Jodoin, M. G., and Ackerman, T. A. (2000, April). *Performance of Mantel-Haenzsel, Simultaneous Item Bias Test and Logistic Regression when the proportion of DIF items is large*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

- Gierl, M. J. (2000). Construct equivalence on translated achievement tests. *Canadian Journal of Education*, 25(4), 280-296. doi:10.2307/1585851
- Gotzmann, A., Wright, K., and Rodden, L. (2006, April). *A comparison of power rates for items favoring the reference and focal group for the Mantel-Haenszel and SIBTEST procedures*. Paper presented at the American Educational Research Association (AERA) in San Francisco, California.
- Gök, B., Kelecioğlu, H. ve Doğan, N. (2010). Değişen madde fonksiyonunu belirlemede Mantel-Haenszel ve Lojistik Regresyon tekniklerinin karşılaştırılması. *Eğitim ve Bilim*, 35(156), 3-16. <http://egitimvebilim.ted.org.tr/index.php/EB/article/view/19/28> adresinden erişilmiştir.
- Greer, T. G. (2004). *Detection of differential item functioning (dif) on the satv: A comparison of four methods: Mantel-Haenszel, Logistic Regression, simultaneous item bias and likelihood ratio test* (Unpublished doctoral dissertation). Retrieved from www.proquest.com (Tez No. 3122349).
- Grover, R. K., and Ercikan, K. (2017). For which boys and which girls are reading assessment items biased against? Detection of differential item functioning in heterogeneous gender populations. *Applied Measurement in Education*, 30(3), 178–195. doi: 10.1080/08957347.2017.1316276.
- Hahne J. (2008). Analyzing position effects within reasoning items using the LLTM for structurally incomplete data. *Psychology Science Quarterly*, 50(3), 379-390. Retrieved from http://journaldatabase.info/articles/analyzing_position_effects_within.html
- Hambleton, R. K., and Traub, R. E. (1974). *The effects of item order on test performance and stress*. *Journal of Experimental Education*, 43(1), 40-46. doi:10.1080/00220973.1974.10806302
- Hambleton, R. K., and Swaminathan, H. (1989). *Item response theory: Principles and applications*. USA: Kluwer Nijhoff Publishing.
- Hambleton, R.K., Swaminathan, H. and Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage.
- Hambleton, R. K. (2006). Good practices for identifying differential item functioning. *Medical Care*, 44(11), 182- 188. doi:10.1097/01.mlr.0000245443.86671.c4
- Herrera, A.-N., and Gómez, J. (2008). Influence of equal or unequal comparison group sample sizes on the detection of differential item functioning using the Mantel-Haenszel and Logistic Regression techniques. *Quality & Quantity: International Journal of Methodology*, 42(6), 739–755. doi: 10.1007/s11135-006-9065-z
- Horgan, J. M. (2012). Programming in R. *WIREs Comp Stat*, 4(1), 75-84. doi: 10.1002/wics.183

- International Test Commission (2005). *International test commission guidelines for test adaptation*. London: Author. Retrieved from https://www.intestcom.org/files/guideline_test_adaptation.pdf
- Jodoin, M. G., and Gierl, M.J. (2001). Evaluating type I error and power rates using an effect size measure with logistic regression procedure or DIF detection. *Applied Measurement in Education*, 14(4), 329-349. doi:10.1207/S15324818AME1404_2
- Johnstone, C. J., Thompson, S. J., Moen, R. E., Bolt, S., and Kato, K. (2005). Analyzing results of large-scale assessments to ensure universal design (Technical Report 41). Retrieved from <http://education.umn.edu/NCEO/OnlinePubs/Technical41.htm>
- Karakoç-Alatlı, B. ve Çokluk-Bökeoğlu, Ö. (2018). Investigation of measurement invariance of literacy tests in the Programme for International Student Assessment (PISA-2012). *Elementary Education Online*, 17(2), 1096-1115. doi: 10.17051/ilkonline.2018.419357
- Kingston, N. M., and Dorans, N. J. (1984). Item location effects and their implications for IRT equating and adaptive testing. *Applied Psychological Measurement*, 8(2), 147–154. doi: 10.1177/014662168400800202
- Kleinke, D. J. (1980). Item order, response location, and examinee sex and handedness on performance on multiple-choice tests. *Journal of Educational Research*, 73(4), 225–229. doi:10.1080/00220671.1980.10885240
- Koyuncu, İ., Aksu, G. ve Kelecioğlu, H. (2018). Mantel-Haenszel, Lojistik Regresyon ve Olabilirlik Oranı değişen madde fonksiyonu inceleme yöntemlerinin farklı yazılımlar kullanılarak karşılaştırılması. *İlköğretim Online*, 17(2), 902-925. doi 10.17051/ilkonline.2018.419339
- Luo, L., Arizmendi, C., and Gates K. M. (2019). Exploratory Factor Analysis (EFA) programs in R. *Structural Equation Modeling: A Multidisciplinary Journal*, 26(5), 819-826. doi: 10.1080/10705511.2019.1615835
- Loehlin, J. C. (1987). *Latent variable models: An introduction to factor, path, and structural analysis*. Hillsdale, NJ: Erlbaum.

- Lyons-Thomas, J., Sandilands, D. D., and Ercikan, K. (2014). Gender differential item functioning in mathematics in four international jurisdictions. *Education and Science*, 39(172), 20-32. Retrieved from <http://egitimvebilim.ted.org.tr/index.php/EB/article/view/2873/625>
- Magis, D., Beland, S., and Raiche, G. (2018). *Collection of methods to detect dichotomous Differential Item Functioning (DIF)*. Retrieved from <https://cran.r-project.org/web/packages/difR/difR.pdf>
- Magis, D., Beland, S., Tuerlinckx, F., and De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods*, 42(3), 847-862. doi:10.3758/BRM.42.3.847
- Mertler, C. A. and Vannatta, R. A. (2005). *Advanced and multivariate statistical methods: Practical application and interpretation*. Los Angeles, LA: Pyrczak.
- Millî Eğitim Bakanlığı (2018). Sınavla öğrenci alacak ortaöğretim kurumlarına ilişkin merkezî sınav başvuru ve uygulama kılavuzu. Ankara. https://www.meb.gov.tr/meb_iys_dosyalar/2019_04/03134315_Kilavuz2019.pdf adresinden erişilmiştir.
- Osadebe, P. U., and Agbure, B. (2018). Assessment of differential item functioning in social studies multiple choice questions in basic education certificate examination. *European Journal of Education Studies*, 4(9), 236-257. doi: 10.5281/zenodo.1301272
- Osterlind, S. J., and Everson, H., T. (2009). *Differential item functioning*. Thousand Oaks, CA: SAGE Publications, Inc.
- Raju, N. S., Laffitte, L. J., and Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology*, 87(3), 527-529. doi:10.1037//0021-9010.87.3.517
- RStudio (2020). *Take control of your R code*. Retrieved from <https://rstudio.com/products/rstudio/>
- Ryan, K. E., and Chiu, S. (2001). An examination of item context effects, DIF, and gender DIF. *Applied Measurement in Education*, 14(1), 73-90. doi:10.1207/S15324818AME1401_06
- Shealy, R., and Stout, W. F. (1993). A model-based standardization approach that separates true bias/ DIF from group ability differences and detects test bias/DTF as well as item bias/DIF. *Psychometrika*, 58(2), 159-194. doi:0033-3123/93/0600-91330500.75/0

- Stark, S., Chernyshenko, O. S., and Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: Toward a unified strategy. *Journal of Applied Psychology*, 91(6), 1292-1306. doi:10.1037/0021-9010.91.6.1292
- Stiglic, G., Watson, R., and Cilar, L. (2019). R you ready? Using the R program for statistical analysis and graphics. *Research in Nursing and Health*, 42(6), 494-499. doi:10.1002/nur.21990
- Stoneberg, B. D. (2004). A study of gender-based and ethnic-based Differential Item Functioning (DIF) in the Spring 2003 Idaho Standards Achievement tests applying the Simultaneous Bias Test (SIBTEST) and the Mantel-Haenszel chi-square Test. (Yayın No. Mart, 2004). Retrieved from <http://learndesign.com/files/eric.ed.gov/fulltext/ED489949.pdf>
- Stout, W., and Roussos, L. (1995). *SIBTEST user manual*. Urbana: University of Illinois.
- Swaminathan, H., and Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27(4), 361-370. <https://doi.org/10.1111/j.1745-3984.1990.tb00754.x>
- Tabachnick, B. G., and Fidell, L. S. (2007). *Using multivariate statistics*. Boston, MA: Pearson.
- Tan, Ş. (2009). Soru sırasının madde güçlüğü ve ayrıcalık gücüne etkisi. *e-Journal of New World Sciences Academy*, 4(2), 486-493. http://www.newwsa.com/download/gecici_makale_dosyaları/NWSA-1091-3-2.pdf adresinden erişilmiştir.
- Thissen, D. (2001) OOADIF v.2.0b: Software for the computation of the statistics involved in item response theory likelihood-ratio tests for differential item functioning. Retrieved from <http://www.unc.edu/~dthissen/dl.html>
- Using Rtools4.0 on Windows (2020). Retrieved from <https://cran.rstudio.com/bin/windows/Rtools/>
- Ünsal-Özberk, E. B. ve Koç, N. (2017). WÇZÖ-IV maddelerinin cinsiyet ve sosyo-ekonomik düzey açısından işlev farklılığının belirlenmesinde kullanılan yöntemlerin karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 8(1), 112-127. doi:10.21031/epod.287577
- Van Buuren, S. (2012). *Flexible imputation of missing data*. Boca Raton, FL: Chapman and Hall/CRC.
- Vandenberg, R. J., and Lance, C. E. (2000). A review and synthesis of the MI literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4-69. doi:10.1177/109442810031002

- Walzebug, A. (2014). Is there a language-based social disadvantage in solving mathematical items? *Learning, Culture and Social Interaction*, 3(2), 159-169. doi:10.1016/j.lcsi.2014.03.002
- Yen, W. M. (1980). The extent, causes and importance of context effects on item parameters for two latent trait models. *Journal of Educational Measurement*, 17(4), 297-311. doi:10.1111/j.1745- 3984.1980.tb00833.x
- Yıldırım, H. (2015). *2012 yılı seviye belirleme sınavı matematik alt testinin madde yanlılığı açısından incelenmesi* (Yayınlanmamış yüksek lisans tezi). <https://tez.yok.gov.tr/UlusalTezMerkezi/>'nden erişilmiştir. (Tez No. 388230).
- Zumbo, B. D., and Thomas, D. R. (1996). *A measure of DIF effect size using logistic regression procedures*. Paper presented at National Board of Medical Examiners, Philadelphia, PA. Retrieved from <http://www.educ.ubc.ca/faculty/zumbo/cv.htm>.
- 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.

Etik Kurul Kararı

Bu araştırma, 01.01.2020 tarihinden önce yapıldığı için etik kurul kararı zorunluluğu taşımamaktadır.



Using R to Detect Differential Item Functioning: Science sub-test of Secondary School Entrance Examination

ARTICLE TYPE	Received Date	Accepted Date	Published Date
Research Article	02.04.2020	09.23.2020	09.26.2020

Betül Alathlı ¹ and Selma Şenel ²
Balıkesir University

Abstract

Differential Item Functioning (DIF) analyses provide critical information about validity of a test. R; an open source software, that comprises all of the DIF detection methods, has an important role in DIF research. Therefore, conducting a guiding study for measurement invariance or DIF analyses by following scientific methods and procedures will be very useful for researchers and practitioners. In this research, it is aimed to illustrate the procedures followed in different DIF detection methods in R, beginning from the installation of the R software to the interpretation of the analysis results, using a sample test (science sub-test of Secondary School Entrance Examination) and data. Four DIF detection methods, which are commonly used in DIF analyses, Mantel-Haenszel, Logistic Regression, SIBTEST and Likelihood Ratio methods are handled in this study. According to the analysis results, no items indicate DIF or indicate negligible DIF.

Keywords: difR, measurement invariance, differential item functioning, item position effect, secondary school entrance examination.

The Ethical Committee Approval: Since this research was conducted before 01.01.2020, it does not require an ethics committee decision.

¹Corresponding author: Assist. Prof., Necatibey Education Faculty, Department of Educational Sciences, e-mail: betulkarakocalatli@gmail.com, <https://orcid.org/0000-0003-2424-5937>

²Assist. Prof., Necatibey Education Faculty, Department of Educational Sciences, e-mail: selmahocuk@gmail.com, <https://orcid.org/0000-0002-5803-0793>

Purpose and Significance

Measurement invariance is a critical indicator for fair decisions obtained from measurement tools. It is defined; individuals who have the same test score from the tests developed to measure a certain latent construct, have the same observed scores at the sub-tests and test items. It is statistically analyzed by Differential Item Functioning (DIF). DIF is defined as the differentiation of the probabilities of achievement of individuals in different subgroups but with the same skill level for an item in the test.

In the literature, numerous DIF detection methods are used, which follow different mathematical processes and use different breakpoints and algorithms. It is recommended that more than one DIF detection method should be used together and they must be compared for investigation of the measurement invariance. However, each DIF method can be applied to different statistical software. As a result of the necessity of using various DIF detection methods, literature reports use of more than one statistical software in one research (Adedoyin, 2010; Akalın, 2014; Çepni, 2011; Gök, Kelecioğlu and Doğan, 2010; Grover and Ercikan, 2017; Karakoç-Alatlı and Çokluk-Bökeoğlu, 2018; Lyons-Thomas, Sandilands and Ercikan, 2014; Stoneberg, 2004; Walzebug, 2014; Yıldırım, 2015).

The software commonly used in DIF analysis are as follows: DIFAS, JMETRIK, EZDIF, Zumbo SPSS Syntax, IRTPRO and IRTLDRIF. R; an open source software, that brings analyses together of all DIF detection methods, has an important role in DIF researches. It is a software that is able to be developed by its users, that is why it is called as *open source*. R provides to make analysis in different fields, and it is widely used due to its free use. Therefore, conducting a guiding study for measurement invariance or DIF analyses by following the scientific procedures, will be very useful for researchers and practitioners. In this research, it is aimed to illustrate the procedure followed in different DIF detection methods in R, beginning from the installation of the R software to the interpretation of the analysis results, through a sample test and data. Four DIF methods, which are commonly used in DIF analyses, Mantel-Haenszel (MH), Logistic Regression (LR) and SIBTEST (Simultaneous Item Bias Test) methods based on Classical Test Theory and Likelihood Ratio (IRT-LR) method based on Item Response Theory are handled in this study.

DIF is examined in the literature for many variables (gender, culture, school type, socio-economic level). *Item position effect* is one of these variables (Avcu, Tunç and Uluman, 2018; Balta and Ömür-Sünbül, 2017; Chiu, 2012; Chiu and Irwin, 2011; Ryan and Chiu, 2001). In order to increase reliability and prevent cheating, especially in achievement tests, different forms (booklet types) are prepared with different item order. In this study, DIF of science items in the Central Exam for Secondary Education Institutions (MEB, 2018) was investigated according to the booklet type. The steps of DIF analysis on R, presentation and interpretation of the analyses results according to different DIF methods are given.

Method

The research is designed as a descriptive survey model (Fraenkel and Wallen, 2006). The research group consists of 2383 8th grade students who took Central Exam for Secondary Education Institutions in 2017-2018 academic year. DIF detection was tested considering two randomly selected booklets (X and Y). The number of students receiving the X booklet is 1208 (50.7%), and the number of students receiving the Y booklet is 1175 (49.3%).

Research data was obtained by permission for research from the General Directorate of Assessment and Examination Services of the Ministry of Education. Since this research was conducted before 01.01.2020, it does not require an ethics committee decision. The 20 science items are handled within the scope of the research. Test items are multiple choice and binary (1-0) scored items.

Initially, data were examined for DIF analysis methods. The normal distribution, extreme and missing values were examined for the suitability of the data for DIF analysis in SPSS. The IRT assumptions; model fit of data, unidimensionality and local independence, was also examined. The data fitted 2-parameter logistic model (2PL). An item (item 7) that was not fit with 2PL was removed from the data set. The analyzes were carried out with 19 science items in X and Y booklets. At this stage, the installation of the R, the conversion of the data set to the appropriate format and the steps for analysis model fit are mentioned respectively.

Step 1 is the installation of R and R Studio. Because R is a freely accessible software, it can be downloaded from the official website (<https://cran.r-project.org/>). Since the downloaded *R Console* has a non-graphical interface; the *R studio* (RStudio, 2020) that offers a user-friendly interface is recommended to be installed. Finally, to continue working on R on Windows, R tools (Using Rtools40 on Windows, 2020) must be installed to run on Windows.

Step 2 is downloading package programs. R provides an environment for analysis for many scopes. Therefore; R has programming packages that contain codes for a certain subject. For example, "*ltm*" (Latent Trait Models under Item Response Theory) package for model data fit reviews or "*difR*" for DIF analyzes are two of them. For package installation, *install.packages("ltm")* code is used. The following *library(ltm)* command is used to call and work on "ltm" package.

Step 3 is editing data, to be used in R. Item responses must be coded as 1-0 with the name of each item (X1, X2... X20) in MS Excel. And the Excel file should be saved as in type *CSV (comma separated)*.

Step 4 is determining the working directory where the analysis files will be saved. In this study, a folder named *DIF* was created in the *D* drive of the computer. The data file named *veri.csv* has also been copied to the folder. Finally, the working directory has been determined with the command *setwd("D: / DIF /")*.

Step 5 is reading from the data file named *veri.csv* that consisted of 20-items. For this, `m1 = read.csv('veri.csv', header = 1, sep = ";") [1: 20]` code should run. `header = 1`, the first line is the header; `sep = ";"` means that the data is comma separated; `[1: 20]` means that the data matrix "m1" has been created for data from column 1 to column 20 in data file. Analysis in the R program is carried out through this data matrix.

Step 6 is analyzing model data fit for Item Response Theory (IRT). Discrete codes should be run for each IRT model. For 1PL model `IRTmodel1 = rasch(m1, IRT.param = TRUE)`, for 2PL model `IRTmodel1 = rasch(m1, IRT.param = TRUE)` and for 3PL model `IRTmodel3 = tpm(m1, type = "latent.trait", IRT.param = TRUE)` codes should be run. Log-likelihood values are obtained for each code execution. However, to examine model fit for each model `item.fit (IRTmodel1); item.fit (IRTmodel2); item.fit (IRTmodel3)` codes should run. The difference between the log likelihood values of the models is taken into consideration in the evaluation of model data fit. As a result of analyses data was fitting with the 2PL model. Item 7 was excluded from the analysis because it was not fit with the 2PL model.

After this stage, DIF analyses should be conducted for the items. The *veri.csv* file is updated by removing item-7 and adding the group variable (G) in which the booklet type is encoded (1-0). The updated data is pulled to R as stated in step 5. Additionally, group variable data in the last column of the *veri.csv* file is taken with `g = read.table('data.csv', header = 1, sep = ";") [, c('G')]` code.

For DIF analysis, "ltm", "lme4", "mirt" packages (as stated in step 2) must be downloaded together with the "difR" package. The analysis can be carried out with the codes related to the DIF method (see Table 6). The analysis is completed by running the code lines in Table 6. The results of the analysis will be saved in the file name specified in the output parameter to the previously specified location (D: / DIF).

Results

In MH method analysis output, Chi square statistics (Stat.) and significance levels (*p-value*) are presented. Accordingly, the results of the chi-square test, none of the items indicate DIF ($p > .05$). The MH odds ratio estimation is given in the *alphaMH* column and the effect size is given in the *deltaMH* (ΔMH) column in the output. The negative ΔMH value of the items means that item provide an advantage for the reference group and positive value means vice versa. There are different classifications of DIF levels. ETS Delta scale is the category ranges determined by Educational Test Service. The ranges of these categories are included in the analysis output as *effect size codes: 0 'A' 1.0 'B' 1.5 'C'*. As a result, according to the MH method, no items identified DIF.

In LR method outputs; statistics (Stat.) and significance levels (*p-value*) are presented. The statistics were not significant for any item ($p > .05$). That means there was no DIF in test items. The expression *Items detected as DIF items: No DIF item detected* in the output file likewise supported the results.

In IRT-LR method outputs, the Likelihood Ratio Statistic is in *Stat.* column and significance levels are presented in the column *p-value*. The items X8, X14, X16 have a significance level below .05. It means these items show DIF. The *Stat.* values show G^2 statistics. G^2 values of items showing DIF were estimated as 4.4213, 4.6874, 4.0893, respectively. Items show negligible DIF according to the G^2 intervals of Greer (2004).

In SIBTEST method output, β , SE , X^2 , df , p means are beta, standard error (standard error-SE), X^2 (Chi-square), degree of freedom and significance level(p). Since all of the p values are greater than .05, that means none of the items indicate DIF. If an item may indicate DIF, "beta" values must be examined. DIF level could be decided according to the critical values specified in the literature (Gotzmann, Wright and Rodden, 2006; Stout and Roussos, 1995).

Discussion and Conclusions

It is anticipated that this study contributes to the researches that study on DIF. Although there are studies about the use of R software (Horgan, 2012; Stiglic, Watson and Cilar, 2019; Luo, Arizmendi and Gates, 2019) or research on detecting DIF in R (Magis, Béland, Tuerlinckx and Boeck, 2010), there is not any study that reports the steps for DIF methods of IRT and CTT.

In the literature, the effect of item order appears to be source for DIF (Avcu et. al., 2018; Balta and Ömür-Sünbül, 2017; Chiu, 2012; Chiu and Irwin, 2011; Ryan and Chiu, 2001). However, there is no research that examines DIF for item orders on large scale tests held in Turkey. According to the results of the analysis, items did not indicate DIF according to the two different booklets used in the "Central Exam for Secondary Education Institutions". In other words, none of the items show DIF among the groups according to the varied order of the items. Accordingly, similar studies can be carried out for other large-scale tests. Similar studies can be conducted with different DIF detection methods.

The Ethical Committee Approval

Since this research was conducted before 01.01.2020, it does not require an ethics committee decision.