

T.C.
BOLU ABANT İZZET BAYSAL ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
EĞİTİM BİLİMLERİ ANABİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME BİLİM DALI

**BİR TESTİN DMF'Lİ MADDE İÇEREN VE DMF'Lİ
MADDEDEN ARINDIRILMIŞ FORMLARINDAN KESTİRİLEN
YETENEK KESTİRİMLERİNİN KARŞILAŞTIRILMASI**

ÖNDER KÂMİL TÜLEK

BOLU-2019

T.C.
BOLU ABANT İZZET BAYSAL ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
EĞİTİM BİLİMLERİ ANABİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME BİLİM DALI

BİR TESTİN DMF'Lİ MADDE İÇEREN VE DMF'Lİ MADDEDEN
ARINDIRILMIŞ FORMLARINDAN KESTİRİLEN
YETENEK KESTİRİMLERİNİN KARŞILAŞTIRILMASI

Yüksek Lisans Tezi

Hazırlayan
Önder Kâmil TÜLEK

Danışman
Doç. Dr. İbrahim Alper KÖSE

BOLU, OCAK-2019

YÜKSEK LİSANS TEZ ONAY FORMU

Önder Kâmil TÜLEK tarafından hazırlanan “**Bir Testin DMF’li Madde İçeren ve DMF’li Maddeden Arındırılmış Formlarından Kestirilen Yetenek Kestirimlerinin Karşılaştırılması**” adlı çalışma Eğitim Bilimleri Anabilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalında Yüksek Lisans Tezi olarak kabul edilmiştir. (30.01.2019)

Akademik Unvan ve Adı Soyadı**İmza**

Üye (Tez Danışmanı) : Doç. Dr. İbrahim Alper KÖSE

Üye : Doç. Dr. Sevilay KİLMEN

Üye : Dr. Öğr. Üyesi Fatma Betül KURNAZ

Eğitim Bilimleri Enstitüsünün Onayı

Prof. Dr. Türkan ARGON
Eğitim Bilimleri Enstitüsü Müdürü

ETİK İLKELERE UYGUNLUK BEYANI

Yüksek Lisans Tezi olarak sunduđum, “**Bir Testin DMF’li Madde İeren ve DMF’li Maddeden Arındırılmıř Formlarından Kestirilen Yetenek Kestirimlerinin Karřılařtırılması**” bařlıklı alıřmanın yazılmasında bilimsel ve etik kurallara uyduđumu, bařkalarının eserlerinden yararlanılması durumunda atıfta bulunulduđunu, kullanılan verilerde herhangi bir tahrifat yapılmadıđını, tezin tamamının ya da bir kısmının bu üniversite veya bařka bir üniversitede bir tez alıřması olarak sunulmadıđını beyan ederim. (30.01.2019)

Önder Kâmil TÜLEK



Ođlum Ömer Alp'e ...

TEŞEKKÜR

Yüksek lisans eğitimim ve tez çalışmam boyunca değerli görüşlerini zaman ve mekan fark etmeden anlayışla benimle paylaşan tez danışmanım ve kıymetli hocam Doç. Dr. İbrahim Alper KÖSE'ye teşekkürü bir borç bilirim.

Akademik gelişmemde önemli katkıları olan hocalarım Prof. Dr. Zekeriya NARTGÜN'e, Prof. Dr. Zeki ARSAL'a, Doç Dr. Sevilay Kilmen'e, tez çalışmam için görüşlerini esirgemeyen Dr. Öğr. Üyesi Önder SÜNBÜL'e, kıymetli yorumlarıyla beni destekleyen değerli hocam Dr. Öğr. Üyesi Fatma Betül KURNAZ'a;

Çalışmam için kapılarını çaldığımda sonuna kadar açan araştırma görevlisi arkadaşlarım Sibel AYDOĞAN ve İbrahim UYSAL'a, dostum Seda ASLAN'a, teknik konularda yardımcı olan mühendis arkadaşım Osman KARAGÖZ'e ve yüksek lisans eğitimi ile tanıma fırsatı bulduğum öğretmen arkadaşım Ali KILIÇ'a;

Hayatım boyunca aldığım her kararda beni destekleyen, yanımda olan annem, babam ve ağabeyime;

Yorgun, yoğun ve umutsuz zamanlarımda beni cesaretlendiren sevgili eşim Nurbanu TÜLEK'e;

Tanıklık etmeye çalıştığımız her anından küçük bir an dahi kaybetmemek için çok uğraştığım canım oğlum Ömer Alp TÜLEK'e sonsuz teşekkür ederim...

İÇİNDEKİLER

ETİK İLKELERE UYGUNLUK BEYANI.....	ii
İTHAF.....	iii
TEŞEKKÜR.....	iv
İÇİNDEKİLER.....	v
TABLolar DİZİNİ.....	vii
ŞEKİLLER DİZİNİ.....	viii
SİMGELER DİZİNİ.....	ix
KISALTMALAR DİZİNİ.....	x
ÖZET.....	xi
ABSTRACT.....	xiii
I. BÖLÜM.....	1
1. Giriş.....	1
1.1. Problem Durumu.....	1
1.2. Problem Cümlesi.....	4
1.2.1. Alt Problemler.....	4
1.3. Araştırmanın Amacı.....	5
1.4. Araştırmanın Önemi.....	5
1.5. Sınırlılıklar.....	6
1.6. Sayılıtlar.....	6
1.7. Tanımlar.....	6
II. BÖLÜM.....	8
2. Kuramsal Çerçeve İlgili Literatür.....	8
2.1. Ölçme.....	8
2.2. Güvenirlik.....	9
2.3. Kullanışlık.....	10
2.4. Geçerlik.....	10
2.5. Yanlılık ve Değişen Madde Fonksiyonu (DMF).....	13
2.5.1. Arındırma (Purification).....	17
2.6. DMF Belirleme Yöntemleri.....	19
2.7. Madde Tepki Kuramı (MTK) ve DMF Belirleme Yöntemleri.....	20
2.7.1. Madde Tepki Kuramı.....	20

2.7.2. Madde Tepki Kuramı Varsayımları.....	22
2.7.3. Madde Karakteristik Fonksiyonu ve Eğrisi.....	22
2.7.4. Madde Tepki Kuramı Modelleri.....	23
2.7.5. Madde ve Test Bilgi Fonksiyonları	26
2.7.6. Madde Tepki Kuramına Dayalı DMF Belirleme Yöntemleri	27
2.8. İlgili Araştırmalar	29
III. BÖLÜM.....	36
3. Yöntem.....	36
3.1. Araştırmanın Modeli.....	36
3.2. Simülasyon Koşulları	36
3.3. Verilerin Üretilmesi	39
3.4. Verilerin Analizi	40
IV. BÖLÜM.....	42
4. Bulgular ve Tartışma.....	42
4.1. Birinci Alt Probleme İlişkin Bulgular ve Yorumlar	45
4.2. İkinci Alt Probleme İlişkin Bulgular ve Yorumlar.....	47
4.3. Üçüncü Alt Probleme İlişkin Bulgular ve Yorumlar.....	50
V.BÖLÜM	54
5. Sonuç ve Öneriler.....	54
5.1. Sonuçlar	54
5.2. Öneriler.....	56
5.2.1. Araştırmacılara Yönelik Öneriler	56
5.2.2. Uygulayıcılara Yönelik Öneriler	57
KAYNAKÇA	58

TABLolar DİZİNİ

Tablo 2.1. KTK'ya dayalı DMF belirleme yöntemleri.....	20
Tablo 3.1. Simülatif veri üretim planı.....	39
Tablo 4.1. Tüm koşullardan elde edilen korelasyon katsayıları.....	42
Tablo 4.2. Madde sayısının farklılaştırılmasıyla elde edilen korelasyon katsayıları	45
Tablo 4.3. Örneklem büyüklüğünün farklılaştırılmasıyla elde edilen korelasyon katsayıları.....	48
Tablo 4.4. DMF oranının farklılaştırılmasıyla elde edilen korelasyon katsayıları	51



ŞEKİLLER DİZİNİ

Şekil 2.1. DMF göstermeyen bir madde karakteristik eğrisi	15
Şekil 2.2. Tek biçimli DMF gösteren madde karakteristik eğrisi.....	16
Şekil 2.3. Tek biçimli olmayan DMF gösteren bir madde karakteristik eğrisi	17
Şekil 2.4. Madde karakteristik eğrisi	23
Şekil 4.1. Tüm koşulların korelasyon katsayıları	43



SİMGELER DİZİNİ

Σ : sigma

θ : theta, yetenek parametresi

θ_1 : 1. theta, DMF içermeyen maddelerden kestirilen yetenek parametresi

θ_2 : 2. theta, DMF içeren maddelerden kestirilen yetenek parametresi

r_s : Spearman Sıra Farkları Korelasyon Katsayısı

z : Fisher z değeri



KISALTMALAR DİZİNİ

DMF	: Değişen Madde Fonksiyonu
MTK	: Madde Tepki Kuramı
KTK	: Klasik Test Kuramı
s.	: Sayfa



ÖZET

BİR TESTİN DMF'Lİ MADDE İÇEREN VE DMF'Lİ MADDEDEN ARINDIRILMIŞ FORMLARINDAN KESTİRİLEN YETENEK KESTİRİMLERİNİN KARŞILAŞTIRILMASI

Tülek, Önder Kâmil

Yüksek Lisans Tezi

Eğitim Bilimleri Anabilim Dalı

Eğitimde Ölçme ve Değerlendirme Bilim Dalı

Tez Danışmanı: Doç. Dr. İbrahim Alper KÖSE

Ocak - 2019, xiv + 63 Sayfa

Bu çalışmada bir testin DMF'li madde içeren ve DMF'li maddeden arındırılmış formlarından kestirilen yetenek kestirimleri karşılaştırılarak bu kestirimler arasında bir ilişki olup olmadığı incelenmiştir.

Araştırma kapsamında araştırmacı tarafından R 3.4.1 paket programı kullanılarak manipüle edilen değişkenlere göre farklı koşullar altında simülasyon çalışmasıyla veri üretimi gerçekleştirilmiştir. Manipüle edilen değişkenler düzeylerine göre örneklem büyüklüğü ($n=1000$ ve $n=2000$), madde sayısı ($k=40$ ve $k=60$) ve DMF oranı ($d=5\%$ ve $d=10\%$) olarak belirlenmiştir. Değişkenlerin çaprazlanması sonucunda sekiz koşulun her birine uygun olacak şekilde DMF'li madde içeren veriler üretilmiştir. Üretilen veriler kullanılarak her koşulun öncelikle DMF'li madde içeren formlarının yetenek kestirimleri (θ_1) gerçekleştirilmiştir. Ardından DMF içeren maddeler testten arındırılarak aynı formlar üzerinden tekrar yetenekler kestirilmiştir (θ_2). Bu yetenek parametreleri arasındaki ilişki Spearman Sıra Farkları Korelasyon Analizi kullanılarak sekiz koşula göre ayrı ayrı hesaplanmıştır. Hesaplanan korelasyon katsayılarının (r_s) tamamının 0'a yakın değerler olduğu görülmüştür ($p<0.01$). Başka bir ifadeyle madde sayısının 40 ve 60, örneklem büyüklüğünün 1000 ve 2000, DMF oranının 5% ve 10% olarak çaprazlandığı sekiz farklı koşulda da testin DMF'li madde içeren formlarından kestirilen yetenek kestirimleri ile DMF'li maddeden arındırılmış formlarından kestirilen

yetenek kestirimleri arasında manidar bir ilişki bulunmamıştır. İlişkinin olmaması da DMF’li maddelerin testten çıkarılmasının bireylerin testteki başarı sıralamalarını değiştirdiğini göstermektedir.

Anahtar Kelimeler: DMF, arındırma, yetenek kestirimi



ABSTRACT**COMPARISON OF THE EXAM TESTS WHICH INCLUDE DIF ITEMS AND WHICH ARE PURIFIED FROM DIF ITEMS**

Tülek, Önder Kâmil

Master's Thesis

Department of Educational Sciences

Measurement and Evaluation in Education

Supervisor: Assoc. Prof. Dr. İbrahim Alper KÖSE

January - 2019, xiv + 63 Pages

This research investigates the exam tests which include DIF items and which are purified from DIF items. While doing this, the ability estimations and purified DIF items are compared to understand whether there is a correlation between the estimations.

The researcher used to R 3.4.1 packet program in order to compare the items and after this situation; according to manipulated factors, we carried out the data production under different circumstances with the help of simulation study. The manipulated factors were determined levels of sample size ($n=1000$ and $n=2000$), test length ($k=40$ and $k=60$) and percentage of DIF ($d=5\%$ and $d=10\%$). After crossing the factors, the data which included DIF items suitable for each of eight conditions. By using the new data each condition's DIF items' ability estimations were carried out. Afterwards, DIF items purified from the tests and later the abilities were estimated.

The correlation between the ability parameters was calculated by using the Spearman's Rank Correlation Coefficient and these parameters were calculated separately according to the eight conditions. After calculations, all of the coefficients of correlations (r_s)' values were almost zero ($p<0.01$). In other words the test length 40 and 60, sample size 1000 and 2000, percentage of DIF 5% and 10%, when we crossed these parameters in different eight conditions, there was no familiar correlation between the tests that include DIF items and tests of that purified from DIF items. Besides, there was no

correlation between the tests thinking the ability estimations; if we exclude DIF items from the tests, the individuals' test ranking changes, too.

Keywords: DIF, purification, estimate of ability



I. BÖLÜM

1. Giriş

Bu bölümde araştırmanın problem durumu, problem cümlesi ile alt problemler, amacı, önemi, sınırlılıklar, sayılılar ve tanımlar yer almaktadır.

1.1. Problem Durumu

Geçmişten günümüze insanın var olduğu bütün alanlarda insanla beraber eğitim de var olmuştur. Eğitim denildiğinde akıllara öncelikle okulların gelmesi, üzerinde tartışılması gereken bir konudur. Çünkü eğitim okul ile sınırlandırılmayacak kadar geniş bir sistemdir. Öyle ki bisiklet sürmeyi bilmeyen bir çocuğun babasıyla çalışmalarından sonra bisikleti sürebilir hâle gelmesi de bir eğitim sonucu olmuştur. Ertürk (1972)'ün eğitim için yaptığı “bireyin davranışlarında kendi yaşantısı yoluyla, kasıtlı olarak istendik değişiklikler oluşturma süreci” tanımından da anlaşılacağı üzere önemli olan bireyde davranış değişikliğinin meydana gelmesidir. Bu davranış değişikliğinin meydana gelmesi belirli etkinlikler, yaşantılar sonucu olur. Yaşantılar sonucunda istenen davranış değişikliklerinin meydana gelip gelmediğinin ya da değişikliğin ne derece gerçekleştiğinin ve istenmeyen davranışların oluşup oluşmadığının belirlenmesi, eğitime katılanların yaşadıkları güçlüklerin saptanması, eğitim programlarının, gerçekleştirilen öğrenme yaşantılarının etkililiğinin izlenmesi ve sürecin aksayan yönlerinin belirlenmesi gerekmektedir. Bütün bunlar eğitimde ölçme ve değerlendirmenin önemine işaret etmektedir (Atılgan, Kan ve Doğan, 2011; Tekin, 2017; Turgut ve Baykul, 2015).

Tekin (2017)'e göre ölçme geniş anlamıyla, “belli bir nesnenin ya da nesnelerin belirli bir özelliğe sahip olup olmadığının, sahipse sahip oluş derecesinin gözlenip gözlem sonuçlarının sembollerle ve özellikle sayı sembolleriyle ifade edilmesidir”. Tanımdan da

anlaşılacağı üzere ölçme için önemli üç kavram üzerinde durulabilir, bunlar: özellik, araç ve sayı sembolidir. Ölçmenin temeli bu kavramlara dayanmaktadır. Başka bir ifadeyle ölçme işleminden söz edilmesi için bu kavramlar kaçınılmazdır. Değerlendirme ise ölçme sonuçlarının bir ölçütle karşılaştırılarak ölçülen özellik hakkında bir yargıya varma, karar verme işlemidir (Turgut ve Baykul, 2015). Eğitim sisteminin vazgeçilmez parçası olan ölçme ve değerlendirme çalışmaları ile eğitim sürecindeki bireylere kazandırılması hedeflenen özelliklerin kazandırılıp kazandırılmadığı ya da ne derece kazandırıldığı gözlemlenebilir ve gözlem sonuçlarına dayalı olarak bireyler hakkında çeşitli kararlar alınabilir. Ancak zaman zaman bireyler için hayati derecede önemli olabilecek bu kararların alınmasının yapısal olarak kabul edilebilir temellere dayanması gerekmektedir. Bu kabul edilebilir temeller, objektif değerlendirmenin gerekliliğine vurgu yapmaktadır. Nesnel değerlendirmeler yapabilmek ise eğitimde ve psikolojide öncelikle ölçme araçlarının kullanılmasıyla mümkündür. Bu ölçme araçlarıyla uzunluk, ağırlık gibi fiziksel özelliklerin kolaylıkla ölçülmesi sağlanabilir. Ancak yetenek, başarı, ilgi, tutum gibi psikolojik özellikleri ölçmek için belirli özelliklere sahip ölçme araçlarının geliştirilmesi gerekmektedir.

Tekin (2017) tarafından ölçme araçları ölçülmek istenen özelliğin içeriğine göre "maksimum yeterlik" ve "tipik davranış" olarak ikiye ayrılmaktadır. Maksimum yeterlik testlerinde bireyin "ne kadar iyi, ne kadar çok" yaptığıyla ilgilenilir ve yüksek puan alması arzu edilirken; tipik davranışların ölçüldüğü ölçme araçlarında ise, bireyin "ne yaptığı" ile ilgilenilir ve elde edilen puanların "iyi" - "kötü" ölçütleri yoktur (Erkuş, 2011). Eğitimde ölçme araçları uygulanarak gerçekleştirilen ölçme işleminden elde edilen sonuçların ölçme amacına uygun olarak kullanılabilmesi için sahip olması gereken bazı nitelikler bulunmaktadır. Bu nitelikler: geçerlik, güvenilirlik ve kullanışlıdır.

Klasik anlamıyla bir ölçme aracının geçerliği, o ölçme aracıyla ölçülmek istenen özelliğe başka bir özellik karıştırılmadan ölçülebilmesi olarak açıklanabilir. Geçmişten günümüze geçerlik üzerine çeşitli tanımlar yapılmış; yapılan bu tanımlarla beraber geçerlik uzun yıllar boyunca kapsam geçerliği, görünüş geçerliği, ölçüt dayanlı geçerlik ve yapı geçerliği olmak üzere dört türde ortaya konmuştur. Ancak son yıllarda geçerlik üzerine güncel tanımlamalar yapılmakta ve yeni bir oluşumdan söz edilebilmektedir.

Daha geniş ve güncel anlamda geçerlik, test puanları üzerine testin hazırlanma amacı doğrultusunda yapılan yorumların hem kuramsal anlamda, hem de toplanan kanıtlarla desteklenme derecesi olarak düşünülebilir (AERA, APA ve NCME, 2014; akt. Kelecioğlu ve Şahin, 2014). Başka bir ifadeyle test puanlarının geçerliğinden söz edebilmek için ya da ölçme aracının ölçmeyi amaçladığı özelliği ne derece ölçebildiğini belirlemek için test hakkında kanıtlar toplanması gerekmektedir.

Ölçme aracının geçerliğine tehdit oluşturan birçok faktör bulunmaktadır. Ölçme aracının kapsamı, araçta yer alan madde sayısı, puanların güvenilirliği, ölçme aracının uzunluğu (madde sayısı), ortalama güçlük düzeyi, sınav süresinin yetersizliği, kopya çekilmesi vb. faktörler geçerliğe tehdit oluşturabilecek faktörlerdir (Turgut ve Baykul, 2015). Ölçme aracının geçerliğine tehdit oluşturan faktörlerden birisi de yanlılıktır (Clauser ve Mazor, 1998). Yanlılık farklı alt gruplarda bulunan bireylere ait test puanlarının ait olduğu gruba bağlı olarak sistematik hata içermesi olarak tanımlanabilir (Camilli ve Shepard, 1994; Zumbo, 1999). Başka bir ifadeyle yanlılık, testin amacı kapsamında olmayan koşullardan veya testteki maddelerin özelliklerinden dolayı; aynı yetenek düzeyinde fakat farklı alt gruplarda yer alan gruplardan birine avantaj sağlanmasıdır (Zumbo, 1999).

Bir testten elde edilen puanların test ile ölçülmek istenen özellik dışında farklı değişkenlerden de etkilenmesi her ne kadar istenmeyen bir durum da olsa pratikte bu kaçınılmazdır. Testi alan bireylerin bulunduğu alt grupların da bu değişkenlerden ne derece etkilendiği önemlidir. Değişkenlerin alt grupları farklı biçimlerde etkilemesi ise yanlılığa sebep olabilmektedir (Atalay, 2010). Bu durum özellikle sonuçlarına bakarak bireyler hakkında çeşitli kararların alındığı geniş ölçekli sınavlar için ayrıca dikkat edilmesini zorunlu hâle getirmiştir. Öyle ki eğitimin birçok alanında, sıralama ya da seçme amaçlı uygulanan sınavlarda alınan kararlar bireyler için hayati olabilmekte ve bu sınavların niteliği alınan kararların doğruluğuna, isabetli ve yerinde olmasına etki etmektedir. Turgut ve Baykul (2015) da isabetsiz kararların alınmasına sebep olarak değerlendirme için toplanan verileri göstermiştir. Bu durumda yanlı maddelerin testten arındırılması bireyler hakkında verilen hayati kararları değiştirmekte midir? Yanlılık üzerine yapılan birçok çalışmada, SBS, TEOG, ÖSS, PISA, ALES, KPSS gibi geniş

ölçekli sınavlarda DMF içeren maddeler tespit edilmiştir (Bakan Kalaycıođlu ve Keleciođlu, 2011; Bařusta, 2013; Ćepni 2011; Demir, 2013; Dođan ve Öğretmen, 2008; Erdem, 2015; Gök, Keleciođlu ve Dođan 2010; Öğretmen 2006; Yıldırım, 2017). Ancak geniş ölçekli bu sınavlarda DMF içeren maddelerin testten çıkarılmasının sonuçlar üzerinde nasıl bir etki oluşturduğuna dair; başka bir ifadeyle DMF’li maddelerin testten çıkarılmasıyla yeniden belirlenen sonuçlara göre bireylerin sınavdaki başarı sıralamalarının etkilenip etkilenmediđine dair çalışmalar sınırlı sayıdadır.

1.2. Problem Cümlesi

Bir testin DMF’li madde içeren ve DMF’li maddeden arındırılmış formlarından kestirilen yetenek kestirimleri madde sayısı, örneklem büyüklüğü ve DMF oranı altında nasıl deđişmektedir?

1.2.1. Alt Problemler

Belirlenen problem cümlesine bađlı olarak bu araştırma için ařađıdaki alt problemlere cevap aranacaktır.

1. Aynı örneklem büyüklüğü ve DMF oranı ile farklı madde sayısı ($k=40$ ve $k=60$) altında bir testin DMF’li madde içeren ve DMF’li maddeden arındırılmış formlarından kestirilen yetenek kestirimleri iliřkisi ne düzeydedir?
2. Aynı madde sayısı ve DMF oranı ile farklı örneklem büyüklüğü ($n=1000$ ve $n=2000$) altında bir testin DMF’li madde içeren ve DMF’li maddeden arındırılmış formlarından kestirilen yetenek kestirimleri iliřkisi ne düzeydedir?

3. Aynı madde sayısı ve örneklem büyüklüğü ile farklı DMF oranı ($d=5\%$ ve $d=10\%$) altında bir testin DMF'li madde içeren ve DMF'li maddeden arındırılmış formlarından kestirilen yetenek kestirimleri ilişkisi ne düzeydedir?

1.3. Araştırmanın Amacı

Araştırmanın amacı bir testin DMF'li madde içeren ve DMF'li maddeden arındırılmış formlarından kestirilen yetenek kestirimlerinin farklı madde sayısı, farklı örneklem büyüklüğü ve farklı DMF oranı koşulları altında karşılaştırmaktır.

1.4. Araştırmanın Önemi

Eğitimde yetenek, başarı gibi psikolojik özellikleri ölçmek için belirli özelliklere sahip testler kullanılmaktadır. Sonuçlarına dayanarak bireyler hakkında önemli kararların alındığı bu testlerin geçerliğinin kanıtlanabilmesi gerekmektedir. Ulusal düzeyde gerçekleştirilen geniş ölçekli ve sıralama gerektiren sınavlarda kullanılan testlerin maddelerinde geçerlik için önemli tehdit oluşturan DMF tespit edildiği çalışmalara sıklıkla rastlanılmaktadır (Bakan Kalaycıoğlu ve Kelecioğlu, 2011; Başusta, 2013; Çepni 2011; Demir, 2013; Doğan ve Öğretmen, 2008; Erdem, 2015; Gök, Kelecioğlu ve Doğan 2010; Öğretmen 2006; Yıldırım, 2017). Ancak DMF tespit edilen maddelerin testten çıkarılmasıyla yeniden hesaplanan sonuçlara göre test alanların başarı sıralamalarında nasıl bir değişiklik oluşturacağına dair bir çalışmaya çoğunlukla rastlanılmamaktadır. Bireyler hakkında hayati kararların alındığı bu sınavlarda belirli bir gruba avantaj sağlayan maddelerin testte bulunmasının bireyler arasında eşitsizliğe ve adaletsizliğe neden olabileceği düşünülmektedir. Bu nedenle bu maddelerin testten arındırılması gerekli olabilmektedir. Peki bu maddelerin testten arındırılarak tekrar sonuç hesaplanmasının bireylerin yetenek kestirimlerinde ve başarı sıralamalarında bir etkisi olacak mıdır? Bu sorulara cevap bulmak amacıyla gerçekleştirilen çalışmadan elde edilen sonuçların alana katkı sağlayacağı, alan araştırmacılarına ve özellikle geniş ölçekli ve sıralama gerektiren sınav hazırlayıcılara yol gösterebileceği düşünülmektedir.

1.5. Sınırlılıklar

Çalışma araştırmacı tarafından yapay olarak üretilen verilerle sınırlıdır. Simülasyon çalışmalarıyla veri üretmek için manipüle edilen değişkenler, madde sayısı ($k=40$ ve $k=60$), örneklem büyüklüğü ($n=1000$ ve $n=2000$) ve DMF oranı, ($d=5\%$ ve $d=10\%$) olarak sınırlı tutulmuştur.

Çalışmada simülasyon çalışmalarıyla veri üretmek için sabit tutulan değişkenler: DMF bulunan maddelerin tek biçimli olması, test sonuçlarından elde edilen verilerin normal dağılım göstermesi, test puanlamasının iki kategorili yapıya uygun olması, model olarak tek boyutlu MTK modeli olan 3 Parametrelili Lojistik Model seçilmesi ile sınırlıdır.

1.6. Sayıtlılar

Simülasyon çalışmalarıyla üretilen verilerin gerçek durumlara benzer veriler olduğu varsayılmıştır. DMF'li madde içeren test formundaki sınırlı sayıda DMF bulunan maddelerin testten arındırılmasının testin kapsam geçerliğine bir tehdit oluşturmadığı varsayılmıştır. DMF'li madde arındırıldıktan sonra yetenek kestirimleri arasındaki ilişkiye madde sayısının azalmasının etki etmediği varsayılmıştır.

1.7. Tanımlar

Madde Tepki Kuramı (MTK): Klasik Test Kuramı (KTK)'na alternatif olarak geliştirilen ve KTK'nın sınırlılıklarını ortadan kaldıran, matematiksel istatistiksel temellere dayanan bir kuramdır.

Geçerlik: Bir ölçme aracının ölçmek istediği değişkeni araya başka değişkenler karıştırmadan ölçebilmesidir.

Değişen Madde Fonksiyonu (DMF): Bir maddenin aynı yetenek düzeyinde ancak farklı gruplarda bulunan bireyler tarafından doğru yanıtlanma olasılığının farklı olmasıdır.

Yanlılık: DMF bulunan bir maddenin belirli bir grubun lehine çalışmasıdır.

Arındırma: DMF tespit edilen maddelerin testten çıkarılarak DMF'li madde bulunmayan test formunun hazırlanmasıdır.



II. BÖLÜM

2. Kuramsal Çerçeve İlgili Literatür

Bu bölümde, araştırmaya ilişkin kuramsal çerçeve ve ilgili araştırmalara yer verilmiştir.

2.1. Ölçme

Eğitim ve psikolojide ölçme genel anlamda bireyler hakkında belirli kararlar vermek amacıyla yapılmaktadır. Çeşitli ölçme araçlarıyla gerçekleştirilen ölçme işlemiyle bireylerden belirli bilgiler elde edilmektedir. Ölçme araçları yoluyla bireylerden elde edilen bu bilgilerin sonuçlarına dayanarak yapılan değerlendirmeler bireyler için hayati olabilmektedir. Öyle ki birçok ülkede olduğu gibi Türkiye’de de özellikle eğitimin her kademesinde bireylerden bilgiler elde etmek için kullanılan ölçme araçlarının belirli niteliklere sahip olması gerekmektedir. Kullanılan bu ölçme araçlarından biri, hatta sonuçlarına bakarak değerlendirmelerin yapıldığı en çok kullanım alanına sahip olanı testtir. Test kavramını daha iyi anlayabilmek için yapılan tanımları inceleyelim.

Baykul (2010) tarafından test, eğitim ve psikolojide bireylerin özelliklerinin belirlenmesi amacıyla kullanılan ölçme aracı olarak tanımlanırken; Türk Dil Kurumu (TDK) tarafından da “bir kimsenin, bir topluluğun doğal veya sonradan kazanılmış yeteneklerini, bilgi ve becerilerini ölçmeye ve anlamaya yarayan sınama” şeklinde açıklanmaktadır. Erkuş (2011)'a göre ise test sözcüğü çoğu zaman sınav, ölçek, envanter yerine de kullanılmakta olup bu yaklaşım doğru değildir; test sözcüğü "sınama" çağrışımlıdır ve özellikle bireyin maksimum yeterliklerini ölçmede kullanılan araçtır. Testi bireylerin belirli özelliklerini ölçmek için düzenlenen ve onu alan herkes için aynı olan soru ya da işlerden oluşan bir ölçme aracı olarak ele alan Tekin (2017), testleri genel anlamda “maksimum yeterlik” ve “tipik davranış” olarak iki şekilde sınıflamaktır.

Maksimum yeterlik testleri ise, bir kişinin kendisine sunulan bir işi belirli bir durumda ne kadar iyi ya da ne kadar doğru yaptığının saptanması amacıyla geliştirilen testlerdir.

Karasar (2016)'a göre, ölçmede istenen niteliklerin belirlenmesi ve bu niteliklerin yerine getirilmesi için gerekli önlemlerin alınması son derece önemlidir; çünkü niteliksiz bir ölçme, yapılan bütün bilimsel çalışmaları değersiz kılabilir. Ölçme sonuçları elde etmede yararlanılan testlerin amacına uygun olarak hizmet edebilmesi için sahip olması gereken bazı yapısal nitelikler vardır. Bunlar: geçerlik, güvenilirlik ve kullanılşıktır. Bu çalışma kapsamında geçerlik kavramı güvenilirlik ve kullanılşıklık üzerine yapılan açıklamaların ardından daha detaylı biçimde ele alınmıştır.

2.2. Güvenirlik

Bir testin amacına uygun olarak kullanabilmesi için sahip olması gereken özelliklerden biri olan güvenilirlik en genel anlamıyla, bir özelliği ölçmek üzere değişik zamanlarda aynı süreçlerin izlenmesi hâlinde benzer sonuçların elde edilmesidir (Karasar, 2016; Tekin 2017). Bir ölçme aracı ile ne kadar az hatalı ölçmeler yapılıyorsa o ölçme aracı o kadar güvenilirirdir.

Güvenirliğin açıklanmasında farklı kavramlar söz konusudur. Alanyazında bu kavramlar karşımıza kararlılık, duyarlılık ve tutarlılık olarak çıkmaktadır (Baykul, 2010; Tekin, 2017; Turgut ve Baykul, 2015). Güvenirlik anlamlarından kararlılık, bir özelliğin bir ölçme aracıyla farklı zamanlarda ölçülmesi durumunda benzer sonuçlara ulaşılmasıdır. Bir ölçme aracının kararlılık anlamında güvenilirliğe sahip olup olmadığı test-tekrar test güvenilirlik belirme yöntemiyle tespit edilebilir. Bir başka güvenilirlik anlamı olan duyarlılık, ölçme aracının birimiyle ilgilidir. Ölçme aracındaki birim küçüldükçe duyarlılık artar. Tutarlılık anlamında güvenilirlik ise, Atılgan ve diğerleri (2011)'ne göre iç tutarlık anlamında ele alınmaktadır, bu güvenilirlik anlamında bir ölçme aracını oluşturan her bir maddenin ölçme aracının bütünüyle tutarlılığı söz konusudur. Yani bir ölçme aracı hangi amaçla hazırlanmışsa, araçtaki bütün maddeler ayrı ayrı o amaca yönelik ve birbirleriyle tutarlı olmalıdır. Bir ölçme aracının tutarlılık anlamında güvenilirliğe sahip olup olmadığı eş değer yarılar yöntemi, KR-20, KR-21 ve Cronbach alfa gibi bazı iç tutarlık güvenilirlik belirleme yöntemleriyle tespit edilebilmektedir. Bunlardan

Cronbach alfa daha çok likert tipi ölçeklerin güvenilirliğini belirlemede kullanılırken KR-20 ve KR-21, 1-0 şeklinde puanlanmaya uygun testlerde kullanılmaktadır.

Yukarıda belirtilen güvenilirlik belirleme yöntemlerinin dışında paralel (eşdeğer) formlar ve puanlayıcı / puanlama güvenilirlikleri de ölçme aracının güvenilirliğini tespit etmede kullanılır. Puanlayıcı / puanlama güvenilirliği özellikle objektif olmayan testlerde (sözlü yoklama, yazılı yoklama gibi) kullanılmaktadır. Burada puanlayıcı güvenilirliğinde birden çok puanlayıcının aynı testi puanlamasındaki korelasyon katsayısı önemli iken, puanlama güvenilirliğinde ise aynı puanlayıcının aynı testi birden çok defa puanlamasında puanlamalar arası korelasyon katsayısı önemlidir.

2.3. Kullanışlık

Ölçme aracının ekonomiklik, uygulanabilirlik ve puanlanabilirlik durumlarıyla ilgili olmasıdır. Tekin (2017)'e göre kullanışlık bir ölçme aracıda güvenilirlik ve geçerlik temel özelliklerinin sağlanmasından sonra düşünülmesi gereken bir özelliktir.

2.4. Geçerlik

Bir ölçme aracıda bulunması gereken yapısal niteliklerden en önemlisi olarak kabul edilen geçerlik kavramı son yıllarda yeni tanımlamalar ile tartışılmaya başlanmıştır. Aslında bu tartışmalar geçerliğin ortaya çıktığı zamanlara kadar dayanmaktadır. Öyle ki geçerlik sınıflamaları ve ele alınış biçimleri başlangıçta ölçüt dayanaklı geçerlik temelinde iken, daha sonra psikolojik yapılarla ilişkilendirilerek yapı geçerliği merkezinde kabul görmeye başlamıştır. Klasik anlamıyla geçerlik bir ölçme aracının ölçmek istediği özelliği, başka özelliklerle karıştırmadan ölçebilmesi olarak açıklanabilir. Klasik tanımı genel bir kabul görmesine rağmen; geçerliğin tanımı, sınıflandırılması, tespit edilmesi, ölçme aracının uygulanmasıyla elde edilen sonuçların yorumlanması üzerinde bir görüş birliğinden söz edilememektedir (Kelecioğlu ve Şahin, 2014). Geçerlik hakkındaki tartışmaları teknolojik ve teorik değişikliklerle açıklayan Zumbo (1999) test

puanlarının geçerliği üzerine yorumlamaların değiştiğini ve güncel tanımlamaların ortaya çıktığını ifade etmektedir. Zumbo (1999)'a geleneksel anlamda geçerlik:

- Ölçülmesi amaçlanan özelliğin ölçülüp ölçülmediği,
- Geçerlik için güvenilirliğin gerekli olduğu ancak yeterli olup olmadığı,
- Ölçme aracının ölçülmek istenen özelliğe uygunluğu,
- Bazı istatistiksel yöntemlerle açıklanabildiği,
- Ölçme işleminin ya geçerli ya da geçersiz olduğu,
- Araştırmacılar için dört geçerlik türünden yalnızca birinin sağlanmasının yeterli olduğu durumlarına odaklanmaktadır.

Geleneksel anlamda geçerlik üzerine yapılan açıklamaların yanlış olmadığı ancak güncel olarak bu tanımlamaların yeni bir oluşumla açıklanması gerektiğini savunan Zumbo (1999); güncel geçerlik tanımlamalarının amacının geleneksel anlamda geçerliğin kavramsal çerçevesini değiştirmeye ve genişletmeye yönelik olduğunu ifade etmektedir. Güncel anlamda geçerliği de şu şekilde ele almaktadır:

- Yapı geçerliği, geçerliğin merkezinde bulunmaktadır ve bu geçerlik türünün sağlanması gerekmektedir.
- Güvenirliğin geçerlik için gerekli olup olmadığı ya da gerekli ancak yeterli olup olmadığı konusunda tartışma vardır.
- Geçerlik artık ölçme aracının bir özelliğinden çok ölçme aracından çıkarılan bir anlamdır.
- Geçerlik bir süreçtir ve geçerli ya da geçerli değil diye bir karar verilemez.
- Geçerlik testin kullanım amacı ve sonuçlarıyla verilecek kararlar açısından önemlidir.

Geçerliğin güncel anlamına ilişki önemli açıklamalar getiren bir diğer araştırmacı Messick (1995), geçerliği birleştirilmiş geçerlik modeli yaklaşımıyla ele almıştır. Kane (2011)'e göre model Loevinger (1957)'in kapsam, yordama ve zamandaş geçerliğinin yapı geçerliği çatısı altında toplanması görüşüne dayanmaktadır ve modele göre yapı geçerliği bir geçerlik türü değildir; geçerlik için kanıt oluşturabilecek bütün yordama, zamandaş ve kapsam geçerliği ile güvenilirlik süreçlerini içerisine alan çatıdır

(Akt. Keleciođlu ve Şahin, 2014). Bu model temelinde çalışmalar yürüten Messick (1995) geçerliđin geniş bir şekilde tanımlanmasının gerekliliđini ifade ederek geçerlik için kanıt toplama ve testin kullanımı ile test puanlarının yorumlanma sonuçlarının önemine vurgu yapmıştır. Başka bir ifadeyle kapsam ve ölçüt dayanaklı geçerlik için yapılan bütün çalışmalar ve süreçler yapı geçerliđi adı altında birleştirilmektedir.

Yapı geçerliđi, test puanlarının kullanımı ve yorumlanması üzerine toplanan kanıtların kapsam ve ölçüt dayanaklı geçerlik için toplanan bütün kanıtlarla birleştirilmesiyle oluşur. Messick (1995)'in bu görüşünü destekleyenler olduđu gibi karşı çıkanlar da bulunmaktadır. Karşı çıkanlara göre birleştirilmiş geçerlik modelinin merkezi olan yapının gözlenemeyen bir deđişken olduđu ve yapıyı ölçebilmek için test puanlarının araç olarak kullanıldıđı söylenebilir. Ancak yapı gözlenemeyen bir deđişken olmasına rağmen; kapsam gözlenebilir ve belirtke tablosuyla ifade edilebilir olduđu için kapsam geçerliđi ve yapı geçerliđi birlikte ele alınmamalı ayrı ayrı deđerlendirilmedir. Öyle ki örtük bir özellik olan yapının açıklanmasıyla, kapsamın da açıklandıđı ya da kapsamın sağlanmasıyla yapının da sağlandıđı düşüncesi her zaman dođru olmayabilir. Başka bir ifadeyle; test puanlarının yapı ya da kapsam geçerliklerinden herhangi birinin sağlanması diđerinin de sağlandıđı anlamına gelmeyebilir.

Keleciođlu ve Şahin (2014)'e göre geçerlik kavramı konusunda halen süren tartışmalar, fikir ayrılıkları olmakla beraber, anlaşılan noktalar da vardır. Bir testin ölçtüđu özelliđe ilişkin kanıt toplama çalışmaları ortak kabul görmüş bir düşüncedir. Bir deđişkeni ölçmek için kullanılan ölçme aracı bir başarı testi de olsa ölçek de olsa doğrudan bir ölçme söz konusu deđildir. Dolayısıyla o ölçme aracının ölçmeyi amaçladıđı deđişkeni ne derecede ölçtüđu hakkında, başka bir ifadeyle ölçme aracından elde edilen sonuçların geçerliđi hakkında kanıt toplanması gerekmektedir. Bu durumda da geçerlik için ne kadar çok kanıt toplanırsa ölçme aracı hakkında o kadar çok bilgi edinilmiş olur. Yani geçerlik kavramının yapı ya da kapsam merkezli ele alınması gerektiđi tartışmaları bir yana duvara astıđımız resmi daha berrak görebilmemiz için geçerlik adına testin ölçmek istediđi özelliđi ölçüp ölçmediđine ilişkin kanıt toplanması süreci en azından uygulamada bir orta yol olarak kabul görecektir. Bu kanıt toplama sürecinde kullanılan

yöntemler ister kapsam, ister ölçüt, ister yapı olsun hepsinin birleşimiyle beraber resim daha net bir hâle gelecektir (Kelecioğlu ve Şahin 2014).

2.5. Yanlılık ve Değişen Madde Fonksiyonu (DMF)

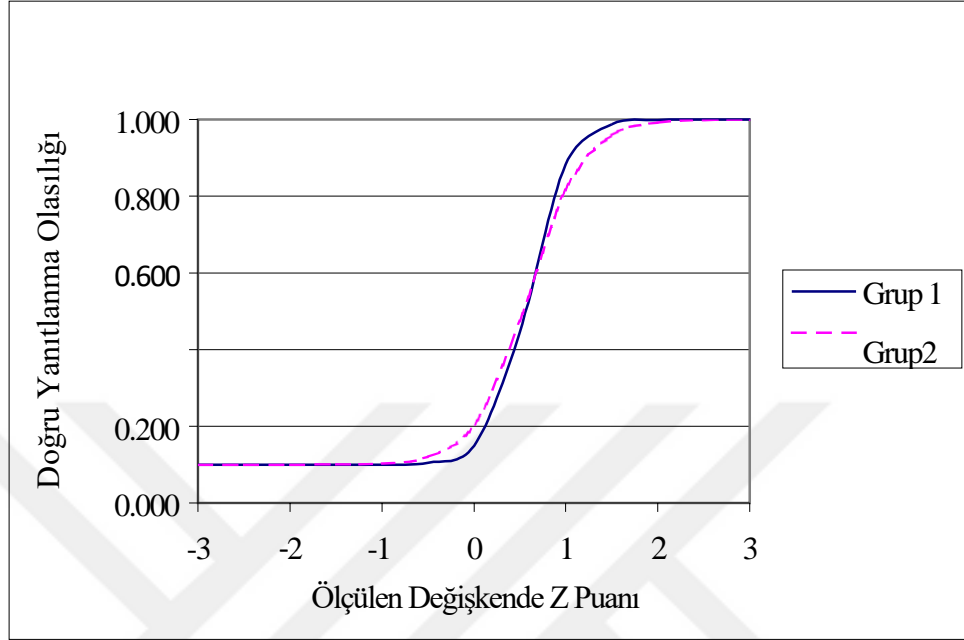
Ölçülmek istenen bir özelliğin gruplar arasında farklılık göstermesinin kaynağı grupların gerçek performanslarındaki farklılık olabileceği gibi belirli ya da belirsiz bir sistematik hata da olabilir (Yurdugül, 2003). Gruplar arasında var olan gerçek farklılık “madde etkisi” olarak adlandırılır (Dorans ve Holland, 1993). Sistematik hatanın sebep olabileceği yanlılık ise “test yanlılığı” ve “madde yanlılığı” olmak üzere iki kategoride incelenmektedir. Test yanlılığı farklı gruplarda yer alan cevaplayıcıların test ortalamalarında ortaya çıkan sistematik bir fark olarak tanımlanabilir (Atılgan ve diğerleri, 2011). Camilli ve Shepard (1994)’a göre test yanlılığı genellikle farklı etnik kökenden gelen gruplar ya da farklı cinsiyete sahip gruplar için söz konusu olabilirken; bu yanlılığın sebepleri test dışı faktörler ve test içindeki maddelerin yapısına dayanabilir. Maddelerin yapısındaki farklılıklar da madde yanlılığına işaret etmektedir.

Madde yanlılığı üzerine alan yazında birçok benzer tanım bulunmakla beraber genel anlamda madde yanlılığı, testin amacı dışındaki koşullardan veya testteki maddelerin karakteristik özelliklerinden dolayı, aynı yetenek düzeyindeki alt gruplarda bir maddenin benzer yanıtlanma olasılığının farklılık göstermesidir (Angoff 1993; Camilli ve Shepard, 1994; Shepard, Camilli ve Williams, 1984; Zumbo, 1999). Geçerliğe önemli bir tehdit oluşturan yanlılık üzerine ilk çalışmalar yapan Alfred Binet’dir. Binet, 1910 yılındaki çalışmasında, düşük sosyoekonomik yapıdan gelen çocukların bazı maddelere verdikleri yanıtların kültürel özelliklerden etkilendiğini fark etmiştir. Bu maddelerin testin amacı dışında kalan özellikleri ölçtüğünü düşündüğü için çalışmasını daha geçerli kılmak amacıyla maddeleri testten çıkarmıştır (Camilli ve Shepard, 1994). Binet ile başlayarak devam eden özellikle 1970’li yıllardan sonraki çalışmaların yanlılığın geçerlik üzerinde ciddi sıkıntı oluşturduğunu ve çoğu konuda verilen kararların isabetsiz olmasına neden olduğunu ortaya koyduğu görülmüştür (Holland ve Wainer, 1993). Bu durum ise son yıllarda araştırmacıların ilgilerini yanlılık kavramına yöneltmelerine ve yanlılık üzerine yapılan kapsamlı çalışmaların artmasına yol açmıştır. Başka bir ifadeyle

yanlılık çalışmaları geniş bir kullanım alanı olan ölçme araçlarının niteliğini tespit etmek adına ölçme kuramcılarının ve uygulayıcılarının odağı hâline gelmiştir (Atalay, 2010). Teknolojik gelişmelerle beraber bilgisayar programlarının da neredeyse bütün analizlere imkân sağlaması; araştırmacıları, çalışmalarını objektif sonuçlarla destekleyebilecek tekniklere yönlendirmiştir.

Madde yanlılığı çalışmalarının amacı, aynı yetenek düzeyindeki bireylerin bulunduğu alt gruplar arasındaki farklılığın ölçülen özellik bakımından gerçekte var olan farklılıktan mı yoksa ölçme işleminden mi kaynaklandığını saptamaktır. Yanlılık çalışmalarında ilk yapılması gereken alt grupların maddeleri yanıtlama yapıları arasında farklılık olup olmadığının belirlenmesidir. Alt grupların maddeleri yanıtlama yapıları arasında farklılık olup olmadığı ise değişen madde fonksiyonu (DMF) analizleri ile mümkündür. Madde yanlılığından farklı bir anlam içeren DMF'ye ilişkin Yurdugül (2003), bir maddeyi doğru yanıtlama olasılığı aynı beceri düzeyindeki iki grup için farklılaşıyorsa maddede DMF bulunduğunu ifade etmiştir. Benzer şekilde Maller (2001) ile Stump, Monahan ve McHorney (2005) DMF içeren bir maddenin aynı yetenek düzeyine sahip iki grupta doğru yanıtlanma olasılığının farklı olduğunu vurgulamışlardır (Akt. Kurnaz, 2006). Hambleton, Swaminathan ve Rogers (1991) ise DMF'yi bir tek test maddesine aynı yetenek düzeyinde fakat cinsiyet, ırk vb. gibi farklı gruplardan gelen iki bireyin doğru cevap verme olasılığının aynı olmaması olarak tanımlamıştır. Görüldüğü üzere DMF tanımlarından çıkarılabilecek ortak kanıya göre, aynı yetenek düzeyinde bulunan bireylerin belirli bir maddeyi benzer şekilde yanıtlanması beklenir. Bu yanıtlama farklılık gösterirse o maddede DMF varlığından söz edilebilir.

Şekil 2.1’de Zumbo (1999)’nun DMF göstermeyen bir maddeye ait madde karakteristik eğrisi gösterilmiştir.



Şekil 2.1. DMF göstermeyen bir madde karakteristik eğrisi (Zumbo, 1999)

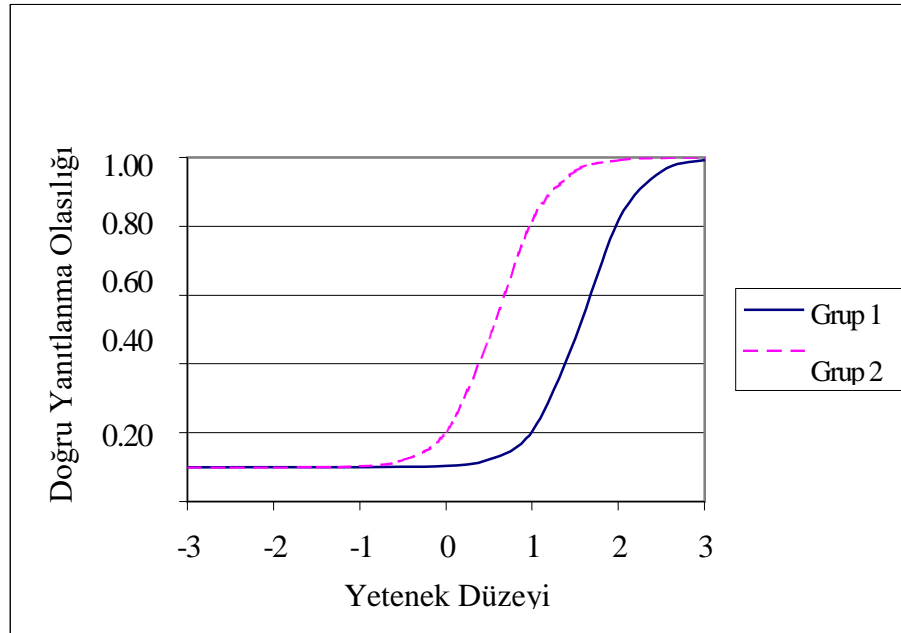
Zumbo (1999) iki grubun madde karakteristik eğrileri birbirine çok benzerse maddenin DMF göstermediğini belirtmektedir. Şekil 2.1.’de Grup 1 ve Grup 2 eğrilerinin oldukça benzer ve iki eğri arasında kalan alanın çok küçük olduğu görülmektedir. Başka bir ifadeyle aynı yetenek düzeyinde fakat farklı alt gruplarda bulunan bu grupların söz konusu maddeyi benzer yanıtladığı anlaşılmaktadır. Bu durumda bu maddeye ilişkin maddenin DMF göstermediği yorumunda bulunabiliriz.

DMF’nin yanlılıktan farklı bir anlam içerdiğini söylemiştik. Bir maddeye yanlıdır diyebilmek için öncelikle o maddenin DMF içermesi gerekmektedir. Ancak bir maddede DMF bulunması, o maddenin kesin olarak yanlı olduğu anlamına gelmez. Başka bir ifadeyle yanlı her madde DMF içermektedir; fakat DMF içeren her madde yanlı değildir. DMF içeren bir maddenin yanlılığına karar verebilmek için DMF’nin olası sebepleri belirlenerek, maddenin ölçmeyi amaçladığı yapı dışında alt gruplardan biri için avantajlı olup olmadığı ile ilgili uzman kanısına başvurulmalıdır (Camilli ve Shepard, 1994; Zumbo 1999). Bu araştırma kapsamında veri setinde yer alan ve DMF gösteren maddeler üzerinde sadece DMF çalışmaları gerçekleştirilmiştir. Bu maddelerin olası bir yanlılık gösterip göstermediği üzerine herhangi bir uzman kanısına başvurulmamıştır.

Başka bir ifadeyle araştırmacı tarafında yanlılık belirleme çalışmalarının ilk adımı olan DMF üzerine çalışmalar gerçekleştirilmiş olup bu adımın devamı niteliğindeki yanlılık belirleme çalışması gerçekleştirilmemiştir.

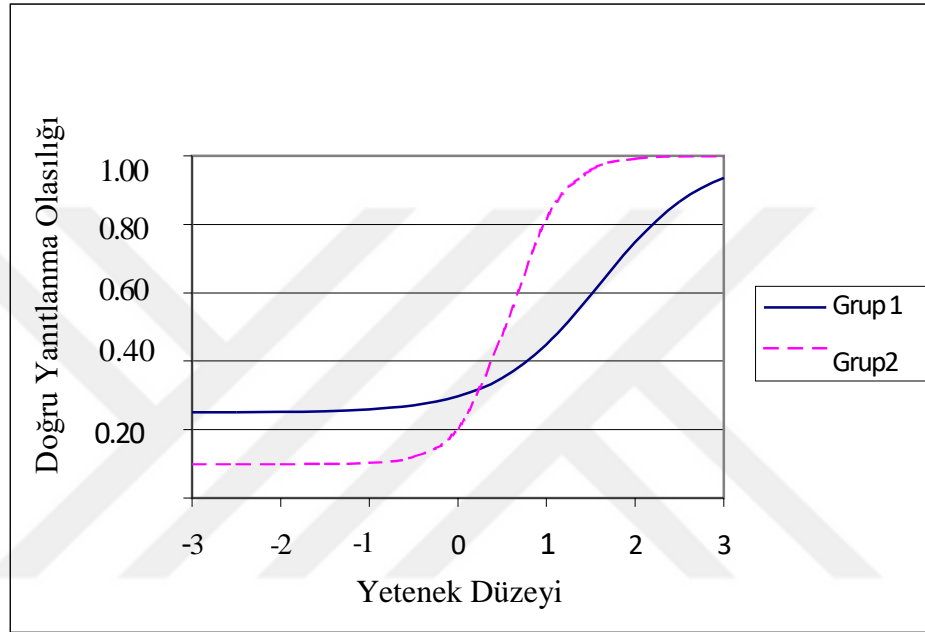
DMF gösteren bir madde tek biçimli (uniform) ve tek biçimli olmayan (non-uniform) olmak üzere iki farklı şekilde ortaya çıkmaktadır. Kamata ve Vaughn (2004)'a göre bir madde aynı yeteneğe sahip iki farklı gruptan birinin tüm yetenek düzeylerinde DMF gösteriyorsa tek biçimli (uniform) DMF, ancak bir grubun örnek olarak yalnız yüksek puan alan bireylerinde DMF gösteriyorsa tek biçimli olmayan (non-uniform) DMF olarak açıklanır. Başka bir ifadeyle tek biçimli DMF gösteren bir madde bütün yetenek düzeylerinde bir grubun lehine işlerken; tek biçimli olmayan DMF gösteren bir madde bir yetenek düzeyine kadar bir grubun, diğer yetenek düzeylerinde ise diğer grubun lehine işlemektedir.

Şekil 2.2.'de tek biçimli DMF; Şekil 2.3.'te ise tek biçimli olmayan DMF gösteren maddelere ilişkin madde karakteristik eğrilerine yer verilmiştir.



Şekil 2.2. Tek biçimli DMF gösteren madde karakteristik eğrisi (Zumbo, 1999)

Şekil 2.2.'de ilgili maddenin "... -3 .. 3 ..." tüm yetenek dağılımında devamlı olarak Grup 2'nin lehine işlediği görülmektedir. Yani Grup 1 ve Grup 2'nin eğrilerinden maddeyi doğru cevaplama olasılığının bütün yetenek düzeylerinde Grup 2'de daha fazla olduğu anlaşılmaktadır. Birbiriyle kesişmeyen iki eğri arasında kalan alan tek biçimli DMF'nin varlığını işaret etmektedir.



Şekil 2.3. Tek biçimli olmayan DMF gösteren bir madde karakteristik eğrisi (Zumbo, 1999)

Şekil 2.3'te iki gruba ait madde karakteristik eğrilerinin yaklaşık 0.3 yetenek düzeyinde kesiştikleri, yani maddenin 0.3 yetenek düzeyine kadar Grup 2 lehine, 0.3 yetenek düzeyinden sonra ise Grup 1 lehine işlediğini görülmektedir. Bu durum maddenin tek biçimli olmayan DMF gösterdiğine işaret etmektedir.

2.5.1. Arındırma (Purification)

Dorans ve Holland (1993)'a göre DMF analizleri için farklı alt gruplardan gelen bireylerin testin ölçmek istediği özellik bakımından benzer olması, yani aynı yetenek düzeyinde eşleştirilmeleri gerekir. Çünkü DMF araştırmaları, testin ölçmek istediği özellik bakımından benzer olan grupların bir maddeye doğru cevap verme olasılıklarının benzer olacağı varsayımına dayanmaktadır. DMF analizlerinde bireylerin özellikle ikili

puanlamaya dayalı testlerden aldıkları toplam puanlar eşleştirme kriteri olarak kullanılmaktadır. Benzer şekilde Clauser, Mazor ve Hambleton (1993) da referans ve odak grupların geçerli bir alt test puanlarının kullanımıyla eşleştirileceğini belirtmişlerdir. Ancak eşleştirme kriteri olarak kullanılacak değişkenin yani ilgili alt test puanlarının hatalardan arınlık derecesi, dikkat edilmesi ve önem verilmesi gereken bir konudur.

Bir testteki maddelere verilen yanıtlar üzerinden hesaplanan ve eşleştirme kriteri olarak kullanılan toplam puanlar, eğer testte DMF içeren madde bulunuyorsa aynı zamanda DMF içeren maddeleri de hesaba katarak elde edilen toplam puanlardır. Bu testten bahsi geçen DMF'li maddelerin çıkarılarak yeniden toplam puan hesaplama işlemi ise arındırma işlemi (purification) olarak adlandırılmaktadır. Kısaca eşleştirme kriterinin arındırılması, toplam puanların hesaplanması sırasında DMF'li maddelerin testten çıkarılması anlamına gelir; böylece gerekli analizler için sadece DMF bulunmayan maddelerin kullanılması sağlanmış olur (Lee ve Geisinger, 2016).

Eşleştirme kriteri olan toplam puanların arındırılmasında iki yaklaşım bulunmaktadır. Bu yaklaşımlardan birincisi Holland ve Thayer (1988) tarafından önerilen iki aşamalı yaklaşım (two-step procedure), diğeri ise yinelemeli (iterative) yaklaşımdır. Bu iki yaklaşım arasındaki temel fark ana analizlerden önce DMF'li maddeleri tespit ederek arındırmak için yapılan ön analiz sayısıdır. İki aşamalı yaklaşımda ilk olarak bir ön analiz ile DMF'li maddeler tespit edilir sonra da bu maddeler testin dışında tutularak toplam puan hesaplanır ve analizler arındırılmış bu toplam puan üzerinde gerçekleştirilir (Holland ve Thayer, 1988). Eğer ön analizler sırasında herhangi bir madde DMF'li olarak tespit edilmeyene kadar analizler tekrarlanıyorsa bu da yinelemeli arındırma yaklaşımıdır (French ve Maller, 2007).

DMF çalışmaları için kullanılan eşleştirme kriterlerinden olan toplam puanın arındırılıp arındırılmaması araştırmacılar tarafından yapılan çalışmalar ve tespitler ile karar verilmesi gereken bir durumdur (Lee ve Geisinger, 2016). Bununla birlikte alanyazında DMF'li maddelerin bulunduğu toplam puanların eşleştirme kriteri olarak kullanılmasının DMF çalışmalarında hatalara yol açabileceği, bu nedenle bu toplam

puanların DMF'li maddelerden arındırılması gerekliliği belirtilmektedir (Holland ve Thayer, 1988; Lee ve Geisinger, 2016, French ve Maller, 2007; Zumbo 1999).

Yukarıda yanlılık üzerine ilk çalışmaların Alfred Binet tarafından 1910 yılında yapıldığı ve Binet'in düşük sosyoekonomik düzeydeki çocukların bazı test maddelerine verdikleri yanıtların kültürel özelliklerden etkilendiğini fark ettiği söylenmişti. Binet de bunun sonucunda bu maddeleri testten çıkarmıştı. Buna göre aslında bilinen ilk yanlılık çalışmasında dahi bir arındırma işlemi yapıldığını söylemek yanlış olmayacaktır.

2.6. DMF Belirleme Yöntemleri

Jodoin ve Gierl (2001)'e göre DMF belirleme yöntemlerinin çoğu tek biçimli DMF'yi belirlemek üzere tasarlanmıştır. Benzer şekilde yine çoğu DMF belirleme yöntemi iki kategorili (1-0) puanlanan testler için yapılandırılmıştır (Zumbo, 1999).

DMF belirleme yöntemleri genel olarak Klasik Test Kuramı (KTK) ve Madde Tepki Kuramına (MTK) dayalı olarak açıklanmaktadır. Genellikle madde ayırt edicilik gücü indeksi ve madde güçlük indeksi gibi madde istatistiklerini karşılaştırma üzerine kurulu KTK'ya dayalı yöntemler bu kuramın getirdiği dezavantajlar da göz önünde bulundurularak çoğu araştırmacı tarafından eleştirilmektedir (Rodney ve Drasgow 1990; akt. Öğretmen, 2006). Başka bir ifadeyle KTK'da hesaplanan madde istatistikleri bir gruptan diğerine değişiklik gösterebildiği için hatalı ölçme sonuçlarına sebep olacağı, bundan dolayı da DMF analizlerinin hatalı yorumlanacağı kabul edilmektedir. Tablo 2.1.'de KTK'ya dayalı DMF belirleme yöntemleri yer almaktadır (Osterlind,1983; Hambleton, Swaminathan ve Rogers, 1991; Camilli ve Shepard, 1994).

Tablo 2.1. KTK'ya dayalı DMF belirleme yöntemleri.

İki kategorili testler için	Çok kategorili testler için
Mantel-Haenszel	Ordinal Lojistik Regresyon (OLR)
Lojistik Regresyon (LR)	Poly-SIBTEST
SIBTEST	
Dönüştürülmüş Madde Güçlüğü İndeksi	
Varyans Analizi	

DMF belirleme yöntemleri üzerine yapılan çalışmalarda yöntemlerin birbirinden farklı sonuçlar verdiği görülmektedir. Osterlind (1983)'e göre bu farklılıkların nedenleri yöntemlere göre kullanılan farklı eşitleme biçimleri, yöntemlerin birbirinden farklı örneklem büyüklükleri gerektirebilmeleri, DMF içeren maddeleri yorumlamada kullanılan farklı kesme noktaları şeklinde belirlenebilir. DMF belirlemede kullanılan yöntemlerin zaman zaman birbiriyle tutarsız sonuçlar vermesinden dolayı DMF çalışmaları için birden çok analiz yönteminin kullanılması önerilebilir.

Bu çalışmada kapsam gereği araştırmacı tarafından KTK'ya dayalı DMF belirleme yöntemlerine yönelik kuramsal açıklamalara gidilmemiştir.

2.7. Madde Tepki Kuramı (MTK) ve DMF Belirleme Yöntemleri

MTK'ya dayalı DMF belirleme yöntemlerini açıklamadan önce MTK'yı tanıtmaya gereğinden dolayı aşağıda MTK hakkında açıklamalara yer verilmiştir.

2.7.1. Madde Tepki Kuramı

Geliştirilen ölçme araçlarının test ve madde analizlerinin KTK'ya dayalı olarak hesaplanması ve elde edilen sonuçların ölçme aracının uygulandığı gruptan gruba değişiklik gösterebilmesi, ölçme araçlarının sonuçları üzerinde bireyleri ve grupları

karşılaştırmayı zorlaştırmıştır. Bu sınırlılık araştırmacıları alternatif kuram arayışına yönlendirmiştir (Hambleton, Swaminathan ve Rogers, 1991). KTK’da karşılaşılan bu sınırlılık ve alternatif kuram arayışı neticesinde Örtük Özellikler Kuramı olarak da bilinen Madde Tepki Kuramı (MTK) ortaya çıkmıştır. Kurama ilişkin ilk çalışmalara 1930’lu yıllarda başlanmış olunmasına rağmen asıl çalışmalar 1950’li yıllardan itibaren süregelmektedir. Lord ve Novick (1968) matematiksel istatistik hesaplamalarına dayanan kuramın temel yapısını yeniden formüle ederek günümüzdeki adıyla Madde Tepki Kuramı’nın temelini oluşturmuşlardır (Akt. Baker ve Kim, 2017). Yetenek kestirimlerinin uygulanan maddelerden bağımsız olarak elde edilmesi kuramın en güçlü yanıdır. MTK’nın dayandığı iki temel yapıya göre bireylerde doğrudan gözlenemeyen örtük özellikler ya da yetenekler (θ), bireylerin test maddelerine verdikleri yanıtlardan kestirilebilir ve bireylerin yetenekleri (θ) ile maddeye karşı gösterdikleri performansları arasındaki ilişki madde karakteristik fonksiyonu adı verilen doğrusal olmayan bir fonksiyon ile açıklanabilir (Hambleton, Swaminathan ve Rogers, 1991).

MTK’ya göre bir madde, o maddeyi yanıtlayacak herhangi bir gruptan bağımsızdır ve bir bireyin yetenek düzeyi, o bireye uygulanan maddelerden bağımsızdır. KTK’ya karşı bu önemli avantajları bulunan, temele aldığı yapıyla KTK’nın sınırlıklarına cevap veren MTK günümüzde sıklıkla kullanılmaktadır. Zhao (2008)’a göre test geliştirme, test sonuçlarından elde edilen puanları eşitleme, madde yanlılıklarını belirleme, bilgisayar ortamında gerçekleştirilen bireysel testlerle yanıtlayıcının yeteneğini kestirme gibi birçok yenilik ve çalışma alanı sağlayan MTK, güçlü varsayımlara dayanması, bu varsayımların karşılanması ve model veri uyumunun sağlanması hâlinde uygulanabilmektedir. (Akt. Köse, 2010). Sözü edilen bu işlemler ileri düzeyde çalışmalar ve araştırmacılar için zorlayıcı uygulamalar gerektirebilmektedir. Uygulamadaki ve hesaplamalardaki bu zorluklar MTK’nın tam anlamıyla KTK’nın yerini alamamasının önemli bir sebebi olarak gösterilmektedir. Öyle ki günümüzde KTK, kolaylıkla uygulanabilir olmasıyla hâlâ fazlasıyla tercih edilmektedir.

2.7.2. Madde Tepki Kuramı Varsayımları

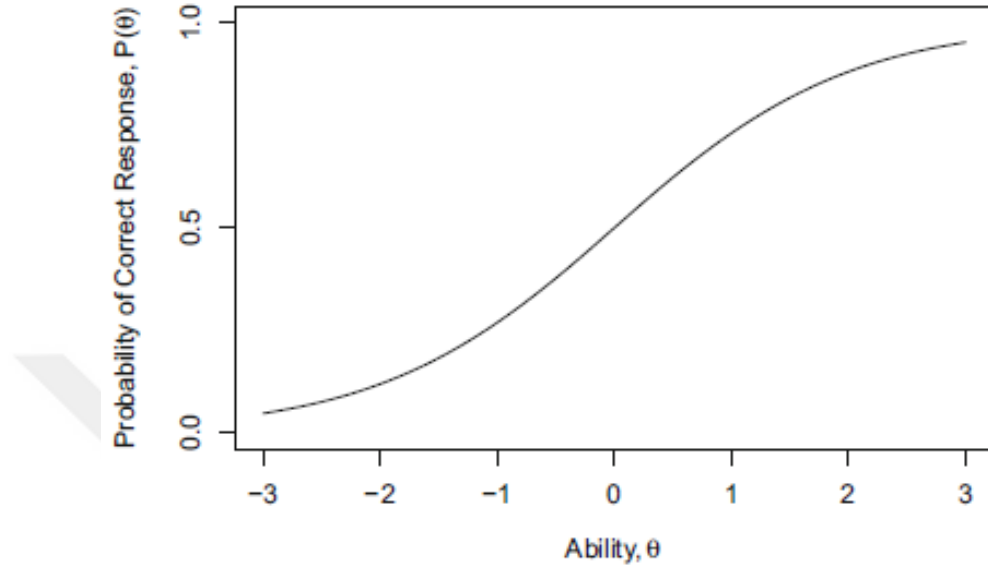
MTK'nın uygulanabilirlik kazanabilmesi için güçlü varsayımları karşılaması gerektiğine yukarıda değinilmişti. Öyle ki varsayımların karşılanmamasıyla model veri uyumuna göre model seçiminde ve sonuçların yorumlanmasında problemler ortaya çıkacaktır.

MTK tek boyutluluk ve yerel bağımsızlık olmak üzere iki temel varsayım üzerine kurulmuştur (Hambleton ve Swaminathan, 1985). MTK modellerinde tek boyutluluk varsayımı bir testte yer alan maddelerin sadece bir örtük özelliği ölçmesi varsayımıdır. Başka bir ifadeyle testin ölçmeyi amaçladığı özellik ile testi alan bireyin performansının kaynaklandığı özellik aynı olmalıdır. Hambleton ve Swaminathan (1985)'a göre bireyin test anındaki performansı kaygı, motivasyon, yanıtlama hızı gibi etmenlerden etkilenmektedir. Bu yüzden uygulamada karşılanması oldukça güç olan bu varsayımın karşılanması testle ölçülen tek bir özellik-yetenek olarak adlandırılan başat (baskın) bir faktör ya da bileşen olmasıyla kabul edilir. Yerel bağımsızlık varsayımına göre ise testi alan bireyin testteki bir maddeye karşı gösterdiği performansın, diğer maddelerden bağımsız olması gerekir. Başka bir ifadeyle aynı yetenek düzeylerinde testteki maddelerin doğru yanıtlanma olasılığı birbirinden bağımsızdır. Bir testin tek boyutluluk varsayımının karşılanmasıyla yerel bağımsızlık varsayımı da karşılanmış demektir (Hambleton, Swaminathan ve Rogers, 1991).

2.7.3. Madde Karakteristik Fonksiyonu ve Eğrisi

Bireyin yeteneği (θ) ile maddeye karşı gösterdiği performans arasındaki ilişkinin madde karakteristik fonksiyonu adı verilen doğrusal olmayan bir fonksiyon ile açıklanabildiği varsayımıdır. Başka bir ifadeyle MTK'nın özü olarak kabul edilen, testi cevaplayanın örtük özelliği (yeterlik/yetenek) ile madde parametreleri arasında kurulan ilişkiye dayalı olasılıklı bir fonksiyondur. Madde karakteristik eğrisi, belirli yetenek düzeyindeki bir bireyin maddeyi doğru yanıtlama olasılığı hakkında bilgi verir (Hambleton ve Swaminathan, 1985). Crocker ve Algina (1986)'ya göre bu eğrinin önemi

KTK'nın madde istatistiklerinin aksine, maddeyi doğru yanıtlama olasılığının bireyde var olan örtük özelliklere nasıl bağlı olduğunu görmemizi sağlamasından kaynaklanmaktadır.



Şekil 2.4. Madde karakteristik eğrisi (Baker ve Kim, 2017).

Şekil 2.4. incelendiğinde testi alan bireyin yeteneği (ability, θ) yatay eksen, maddeyi doğru yanıtlama olasılığı (probability of correct response, $P(\theta)$) ise dikey eksende görülmektedir. MTK'nın madde parametreleri ayırıcılık (a), güçlük (b) ve şans parametresi (c) olmak üzere üç parametre olarak incelenmektedir. Bu üç parametre eğrinin şeklini belirler. Öyle ki eğrinin eğiminin artması, ayırıcılığın arttığına; eğrinin sağa doğru yatması da güçlüğü arttığına işarettir. Şans parametresi ise eğrinin orijinden başlamamasına neden olmaktadır (Baker ve Kim, 2017).

2.7.4. Madde Tepki Kuramı Modelleri

MTK modelleri alanyazında karşımıza farklı sınıflandırmalarla çıkabilmektedir. Genel anlamda MTK modelleri boyutluluğuna göre, maddeye verilen cevapların iki ya da çok kategorili olmasına göre ve doğrusallık yapısına göre sınıflandırılır. Bu çalışmada alanyazında en sık kullanılan, tek boyutlu ve iki kategorili modeller olan, kestirilen parametre sayısına göre bir parametrelili, iki parametrelili ve üç parametrelili lojistik modeller ele alınarak açıklamalar getirilmiştir (Baker ve Kim, 2017; Embretson ve Reise 2000; Hambleton ve Swaminathan, 1985; Hambleton, Swaminathan ve Rogers, 1991).

Bir parametrelili lojistik model (1PLM): Rasch Modeli olarak da bu model MTK'nın en çok kullanılan modellerindedir. Modelde ayırt edicilik parametresinin (a) bütün maddeler için aynı olduğu ve şans parametresinin (c) ihmal edilebileceği varsayılırken yetenek kestirimleri sadece güçlük parametresine (b) göre yapılır. Modelde bir maddenin doğru yanıtlanma olasılığı şu matematiksel ifade ile belirlenir:

- $P_i(\theta)$: θ yetenek düzeyinde bir bireyin i maddesini doğru yanıtlama olasılığı,
 b_i : i maddesinin güçlük parametresi,
 n : testteki madde sayısı,
 θ : yetenek parametresi
 e : 2,718 doğal logaritma tabanı olmak üzere;

$$P_i(\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}} \quad i = 1, 2, 3, \dots, n.$$

İki parametrelili lojistik model (2PLM): Bir parametrelili lojistik modele a parametresi yani ayırt ediciliğin eklenmiş hâlidir. Bu iki parametre madde karakteristik eğrisini belirler. Madde karakteristik eğrisi üzerinde güçlük parametresi, maddenin doğru yanıtlanma olasılığının 0.50 olduğu noktadaki yetenek düzeyi (θ) değeridir. Eğrinin yön değiştirdiği noktadaki eğimi ise ayırtıcılık parametresini ifade eder. Modelde bir maddenin doğru yanıtlanma olasılığı şu matematiksel ifade ile belirlenir:

- $P_i(\theta_j)$: θ yetenek düzeyinde j bireyinin i maddesini doğru yanıtlama olasılığı,
 D : 1,702 ölçek sabiti,
 e : 2,718 doğal logaritma tabanı,
 a_i : i maddesinin ayırıcılık parametresi,
 b_i : i maddesinin güçlük parametresi,
 θ_j : j bireyinin yetenek parametresi
 x_i : i maddesine verilen tepki (1-0) olmak üzere;

$$P_i(\theta_j) = P(x_i=1 \mid a_i, b_i, \theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}}$$

Üç parametrelili lojistik model (3PLM): İki parametrelili modele şans parametresinin (c) dâhil olduğu modeldir. Bu parametre maddeyi tahminle doğru yanıtlama olasılığıdır. Madde karakteristik eğrisi şans başarısından dolayı 0'dan başlamaz, eğrinin y eksenine değen noktası şans parametresidir. Modelde bir maddenin doğru yanıtlanma olasılığı şu matematiksel ifade ile belirlenir:

- $P_i(\theta_j)$: θ yetenek düzeyinde j bireyinin i maddesini doğru yanıtlama olasılığı,
 D : 1,702 ölçek sabiti,
 e : 2,718 doğal logaritma tabanı,
 a_i : i maddesinin ayırıcılık parametresi,
 b_i : i maddesinin güçlük parametresi,
 θ_j : j bireyinin yetenek parametresi,
 c_i : i maddesinin şans parametresi,
 x_i : i maddesine verilen tepki (1-0) olmak üzere;

$$P_i(\theta_j) = P(x_i=1 \mid a_i, b_i, c_i, \theta_j) = \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}}$$

2.7.5. Madde ve Test Bilgi Fonksiyonları

Madde ve test bilgi fonksiyonları ölçmenin standart hatasıyla ters orantılı olup KTK'nın güvenilirlik kavramının karşılığı olarak MTK'da karşımıza çıkmaktadır. Ancak iki kuramın özellikleri arasındaki önemli ayrımı unutmamak gerekir; madde ve test bilgi fonksiyonları, KTK'da kullanılan güvenilirlik katsayısının aksine testin uygulandığı gruptan bağımsızdır (Hambleton ve Swaminathan, 1985; Hambleton, Swaminathan ve Rogers, 1991). Madde bilgi fonksiyonu için kullanılabilir eşitlik:

- $I_i(\theta)$: θ yeteneğinde i. maddenin bilgi fonksiyonu,
 $P_i(\theta)$: θ yeteneğine bağlı madde tepki fonksiyonu,
 $P_i'(\theta)$: $P_i(\theta)$ 'nin birinci türevi,
 $Q(i)(\theta)$: $1 - P_i(\theta)$ olmak üzere;

$$I_i(\theta) = \frac{[P_i'(\theta)]^2}{P_i(\theta) \cdot Q(i)(\theta)} \text{ 'dir.}$$

Üç parametrelili lojistik model için kullanılan maddenin doğru yanıtlanma olasılığının matematiksel ifadesindeki $P_i(\theta)$ 'nin açılımı bu eşitlikte yerine konarak madde bilgi fonksiyonu için daha açık bir eşitlik elde edilebilir:

$$I_i(\theta) = \frac{2,89 a_i^2 (1 - c_i)}{[c_i + e^{1.7 a_i (\theta - b_i)}][1 + e^{-1.7 a_i (\theta - b_i)}]^2}$$

Bu eşitlikten ayırıcılık (a), güçlük (b) ve şans (c) parametrelerinin madde bilgi fonksiyonuna nasıl etki ettiği anlaşılmaktadır. Eşitlik incelendiğinde, ayırıcılık parametresinin (a) artmasının, güçlük parametresinin θ yetenek düzeyine yaklaşmasının ya da şans parametresinin ise (c) 0'a yaklaşmasının maddenin verdiği bilgiyi arttırdığı açıkça görülmektedir.

θ yetenek düzeyinde madde bilgi fonksiyonlarının toplamı, test bilgi fonksiyonuna eşittir. Başka bir ifadeyle testin sağladığı bilgi, maddelerin her birinin

sağladığı bilgilerin toplamıdır. Yani her maddenin birbirinden bağımsız olarak test bilgi fonksiyonuna etkisi vardır. Bu eşitlik şu şekilde gösterilir.

$$I(\theta) = \sum_{i=1}^n I_i(\theta)$$

MTK’da maddeleri ve testleri tanımlamak, testlerin karşılaştırılmasını sağlamak ve test maddelerinin seçiminde kullanmak için madde ve test bilgi fonksiyonları kullanılır.

2.7.6. Madde Tepki Kuramına Dayalı DMF Belirleme Yöntemleri

MTK’ya dayalı DMF belirleme yöntemleri genel anlamda farklı alt gurupların maddeye gösterdiği tepkilerden elde edilen madde karakteristik eğrilerinin karşılaştırılmasıyla gerçekleştirilmektedir (Hambleton ve Swaminathan, 1985; Hambleton, Swaminathan ve Rogers, 1991). Madde karakteristik eğrileri arasındaki alan yöntemi, madde parametreleri farklılığı yöntemi ve Madde Tepki Kuramı Olabilirlik Oran (MTK-OO) yöntemi MTK’ya dayalı DMF belirleme yöntemleri olarak incelemektedir.

Madde karakteristik eğrileri arasındaki alan yöntemi: Bu yöntemde madde karakteristik eğrilerinin karşılaştırılması için Raju (1988) tarafından Raju’nun alan ölçümü adıyla ortaya çıkarılan bir çalışma gerçekleştirilmektedir; bu çalışmada grupların madde karakteristik eğrilerinin çakışıp çakışmadığı ya da eğriler arasında ölçülen bir alan değerinin bulunup bulunmadığı araştırılır. Maddede DMF bulunmaması için eğrilerin çakışması yani eğriler arasında ölçülen alan değerinin sıfır olması gerekir (Akt. Başusta, 2013). Başka bir ifadeyle aynı yetenek seviyesinde aynı maddeyi yanıtlayan iki farklı grubun aynı düzlem üzerinde karşılaştırılan madde karakteristik eğrileri farklılık gösteriyorsa maddede DMF bulunuyor demektir. Yani aynı yetenek (θ) seviyesinde bulunan grupların maddeyi doğru yanıtlama olasılıkları $P(\theta)$ farklılaşmıştır.

Madde parametreleri farklılığı yöntemi: Testi alan iki grubun odak grup ve referans grup olarak ayrıldığı bu yöntemde grupların kestirilen madde parametreleri karşılaştırılır. Embretson ve Reise (2000)'a göre madde parametrelerinin odak ve referans gruplarında farklılaşması, bu grupların maddeyi doğru yanıtlama olasılığının farklılaştığını yani maddede DMF bulunduğunu göstermektedir.

Madde Tepki Kuramı Olabilirlik Oran (MTK-OO) yöntemi: Thissen, Steinberg ve Wainer (1993) tarafından geliştirilen bu yöntemde bireylerin eşleştirilmesinde MTK'ya dayalı olarak kestirilen yetenek kestirimleri kullanılır. IRTL RDIF yazılımı ile MTK-OO analizleri gerçekleştirilebilmektedir. Thissen, Steinberg ve Wainer (1993)'e göre yöntemde gerçekleştirilen çalışma sınırlı ve serbest model olarak tanımlanan iki modelin, olabilirlik oran farklarının anlamlılığının test edilmesidir. Daha açık bir ifadeyle bu yöntem ile yapılan DMF analizinde öncelikle iki grup için bir genelleştirilmiş model kurulur, bu modelde bir veya daha fazla madde parametresinin kestirimi serbest bırakılmıştır. Ardından iki grup için madde parametrelerini aynı olarak kısıtlayan sınırlandırılmış model kurulur. Serbest model, sınırlandırılmış modellerin model veri uyumu belirlenmesinde kullanılan $-2\log$ (olabilirlik oranı) değerleri karşılaştırılarak, modellerin uyumları arasındaki fark G^2 istatistiği olarak raporlanır. Hesaplanan G^2 istatistiği kay kare olarak dağılır ve bu değer tablo değerinden büyükse, yani manidarsa, iki gruptan kestirilen madde parametrelerinin farklı olduğu ve DMF bulunduğu kabul edilir ve sınırlı ve serbest bırakılan modelin parametre sayısının farkı serbestlik derecesi olarak alınır. G^2 değeri manidar değilse serbest bırakılan modeldeki parametrelerin hiçbirinin 0'dan farklı olmadığına karar verilir (Çepni, 2001; Deveci, 2008).

Thissen, Steinberg ve Wainer (1993)'e göre MTK-OO analizi ile DMF çalışmaları yapılırken karşılaştırılan gruplardan kestirilen madde parametrelerini aynı ölçek üzerinde eşitlemek amacıyla iç kriter olarak testin kendi maddelerinden oluşan ortak (ancor-connection) maddeler kullanılır. Sınırlandırılmış modelde iki grup arasında gruplardan kestirilen madde parametrelerinin değişmediği varsayılarak, serbest bırakılan modelde ortak maddeler dışında kalan maddeler için DMF incelenir ve iki modelin olabilirlik oran farklarının manidarlığı test edilir. Ortak maddeler DMF bulunmayan maddeler olarak kabul edilir ve DMF analizlerine dahil edilmez (Kim ve Cohen, 2002).

DMF belirleme çalışmalarının MTK'ya dayalı yapılmasının KTK'ya dayalı yapılmasına karşı avantajları vardır. MTK'da madde parametrelerinin kestirilmesinin uygulanan gruptan bağımsız olarak yapılması ve farklı alt gruplarda değişen maddelerinin daha kesin tespit edilebilmesi bu avantajlardandır. Ayrıca madde parametrelerinin madde karakteristik eğrisinde verilmesinden dolayı DMF bulunan maddelerin kolay anlaşılması MTK'ya dayalı DMF çalışmalarının bir başka üstünlüğüdür (Camilli ve Shepard, 1994). Günümüzde MTK'ya dayalı yöntemler kapsamlı ve daha anlaşılır DMF çalışmaları yapılmasına olanak sağlamaktadır. Narayanan ve Swaminathan (1994)'a göre Madde Tepki Kuramına dayalı yöntemler teorik olarak güçlü yöntemler olmakla birlikte, büyük örneklemeler gerektirir. DMF çalışmalarında ise pratikte bu koşulu karşılamak zordur.

2.8. İlgili Araştırmalar

Bu bölümde başarı testleri üzerine gerçekleştirilen DMF belirleme, simülasyon verileriyle gerçekleştirilen DMF ve eşleştirme kriteri olan toplam puanı DMF'li maddelerden arındırma ile ilgili yurt içinde ve yurt dışında yapılan araştırmalar incelenerek özetlenmiştir.

Roussos ve Stout (1996a) yaptıkları iki aşamalı simülasyon çalışmasıyla DMF belirlenmesinde kullanılan SIBTEST ve Mantel-Haenszel yöntemlerinin I. tip hata performanslarını araştırmışlardır. İlk aşamada yetenek dağılımları benzer ve küçük örneklemelerle simülasyon çalışması gerçekleştirilmiş ve her iki yöntemde de I. tip hata oranlarının düşük olduğu gözlenmiştir. Yer ve grupların sistematik bir şekilde çeşitlendirildiği büyük örneklemeler oluşturularak gerçekleştirilen ikinci çalışmada ise sınava girenlerin yetenek dağılımlarında farklılıklar oluştuğunda iki yöntemde de belirli maddeler için I. tip hata oranlarının arttığı belirlenmiştir. Ayrıca SIBTEST yönteminin hata oranının MH yönteminin hata oranından daha düşük olduğu tespit edilmiştir.

Gierl, Khaliq ve Boughton (1999), Matematik ve Fen bilgisi alanlarına ait bazı başarı testlerinde MH, SIBTEST ve LR olmak üzere üç yöntemi karşılaştırarak cinsiyet değişkenini incelemek üzere DMF analizi yapmışlardır. Elde edilen bulgular

doğrultusunda matematik alt testindeki 469 maddenin MH yöntemine göre % 7.3'ünde, SIBTEST yöntemine göre % 8.5'inde ve LR yöntemine göre ise %10.5'inde DMF'li madde belirlenmiştir. Fen bilgisi alt testinde de 428 maddenin MH yöntemine göre % 8.4'ünde, SIBTEST yöntemine göre % 14.7'sinde ve LR yöntemine göre de %17.3'sinde TB, DMF'li madde tespit edilmiştir. Fen bilgisi alt testinde daha fazla sayıda maddenin DMF'li olarak belirlendiği görülmüştür.

Öğretmen (2006) araştırmasında 2001 PIRLS testinin psikometrik özelliklerini Türkiye ve Amerika Birleşik Devletleri örneklemeleri bağlamında karşılaştırmalı olarak incelemiştir. İki aşamalı olarak gerçekleştirilen araştırmanın birinci aşamasında PIRLS'te uygulanan okuma parçalarına ait testlerin ölçmeyi amaçladığı psikolojik yapıların kültürlere göre eş değer olup olmadığı çok örneklemlili faktör analizi ile ortaya konmuştur. Bu analiz sonucunda test maddelerine ait faktör yük ve hata varyansı değerlerinin kaynak kültürden ve hedef kültürden elde edilen ölçümler arasında manidar bir farklılıkta olduğu tespit edilmiştir. Araştırmanın ikinci aşamasında test maddelerinin kültürlere göre DMF içerip içermediği MTK'ya dayalı parametre karşılaştırma ve olabilirlik oranı testi model karşılaştırma yöntemleri ile incelenmiştir. Bu yöntemlere göre yapılan DMF analizleri sonucunda her iki yöntemine göre test maddelerinin çoğunun kültürlere göre DMF içerdiği gözlemlenmiştir.

French ve Maller (2007) yaptıkları simülasyon çalışmasında, DMF tespiti için LR yöntemini kullanmışlardır. Örneklem büyüklüğü, DMF yüzdesi ve yetenek farklılıklarının manipüle edildiği çalışmanın sonucuna göre genel olarak güç ve Tip I hata oranlarında önemli bir ölçüde düzelme olmamasına rağmen, arındırmanın belirli koşullar altında yararlarının bulunduğu tespit edilmiştir. Ayrıca DMF çalışmalarındaki yanlışları kontrol etmek için arındırma işlemi önerilmiştir.

Doğan ve Öğretmen (2008) tarafından yapılan çalışmada, DMF belirleme tekniklerinden ki kare, Mantel-Haenszel ve LR tekniklerini karşılaştırılıp uygulamada ortaya çıkan benzerlik veya farklılıklar incelenmiştir. Çalışma, 2003 yılında Ortaöğretim Kurumları Seçme ve Yerleştirme Sınavı'na (OKÖSYS) katılan yaklaşık 600.000 öğrenci arasından yansız olarak seçilen 3345 öğrenciden oluşan bir örneklem üzerinde

gerçekleştirilmiştir. Çalışmanın verilerini, OKÖSYS fen bilgisi alt testine öğrencilerin verdiği yanıtlar oluşturmuştur. DMF analizleri yalnızca cinsiyet değişkeni üzerinden yürütülmüştür. Araştırmanın sonuçları, söz konusu tekniklerin bazı açılardan birbirlerine göre benzerlik ve farklılıklar sağladığını göstermiştir.

Awuor (2008) yapay veriler üzerinde gerçekleştirdiği araştırmada DMF belirlemede kullanılan SIBTEST ve Mantel-Haenszel (MH) yöntemlerine farklı örneklem büyüklüklerinin etkisini belirlemeyi amaçlamıştır. Veriler WinGen2 programı ile 2 PLM'e uygun olarak üretilmiştir. Çalışma sonucunda oluşturulan farklı örneklem büyüklüklerinde MH yönteminin SIBTEST yönteminden daha iyi I. tip hata kontrolü sağladığı tespit edilmiştir.

Gök, Kelecioğlu ve Doğan (2010) tarafından yapılan araştırmada, 2005 yılında uygulanan Ortaöğretim Kurumları Sınavı (OKS)'nın matematik ve fen bilgisi alt testlerinde, maddelerin değişen madde fonksiyonu içerip içermediği cinsiyet ve okul türü değişkenlerine göre incelenmiştir. Analizler için Mantel-Haenszel (MH) ve lojistik regresyon (LR) teknikleri kullanılarak karşılaştırılmıştır. Elde edilen bulgulara göre, MH ve LR teknikleri arasında matematik ve fen bilgisi alt testlerinde cinsiyet ve okul türü değişkenlerine göre değişen madde fonksiyonu içeren maddeleri belirlemede genel anlamda düşük düzeyde bir uyum görülmüştür.

Bakan Kalaycıoğlu ve Kelecioğlu (2011), 2005 yılı Öğrenci Seçme Sınavı'nda (ÖSS) yer alan maddelerin cinsiyet değişkenine göre DMF içerip içermediğini ve DMF içeren maddelerin uzman görüşüne dayalı olarak yanlı olup olmadığını araştırmışlardır. DMF belirleme çalışmaları Mantel-Haenszel (MH) ve Lojistik Regresyon (LR) teknikleri ile gerçekleştirilmiştir. DMF analizleri için 2005 yılında ÖSS'ye giren 599,330 lise son sınıf öğrencisinin her biri 45 maddeden oluşan Türkçe, Sosyal bilimler, Matematik ve Fen Bilimleri alt testlerine verdikleri cevaplar kullanılmıştır. Araştırmadan elde edilen sonuçlar doğrultusunda, cinsiyet değişkenine göre Türkçe alt testini oluşturan maddelerin DMF içermediği, Sosyal bilimler alt testinde yedi, Matematik ve Fen bilimleri alt testlerinde de üç maddenin DMF içerdiği ve bu maddelerden fen bilimleri testinde yer alan bir maddenin yanlılık gösterdiği belirlenmiştir.

Çepni (2011) araştırmasında Akademik Personel ve Lisansüstü Eğitimi Giriş Sınavı (ALES) sayısal yetenek testlerinde cinsiyete ve adayların mezun olduğu/olacağı bölüme göre değişen madde fonksiyonu (DMF) ve değişen madde grubu fonksiyonu (DMGF) analizleri gerçekleştirmiştir. DMF analizleri için Mantel Haenszel (MH), lojistik regresyon (LR), SIBTEST, MTK-Olabilirlik Oranı ve BILOG-MG DMF Algoritması; DMGF analizleri için de SIBTEST yöntemi kullanılmıştır. Araştırmanın verilerini, ALES 2008 Sonbahar uygulamasında adayların sayısal yetenek testlerine verdikleri yanıtlar oluşturmuştur. Cinsiyete değişkenine göre DMF analizlerinde Sayısal 1 Testinde üç maddenin erkek öğrencilere, dört maddenin ise kız öğrencilere avantaj sağladığı görülmüştür. Sayısal 2 Testinde de bir madde erkek öğrencilere avantaj sağlarken, 3 madde de kız öğrencilere avantaj sağlamıştır. Bölüm değişkenine göre DMF analizleri gerçekleştirmek için adayların mezun oldukları bölümler sayısal, sözel ve eşit ağırlık olmak üzere üç grupta sınıflandırılmıştır. Sayısal 1 testi ile Sözel bölümler karşılaştırıldığında belirgin farkların ortaya çıktığı gözlemlenmiştir. 5 madde sayısal bölüm adaylarına avantaj sağlarken, 4 madde ise sözel bölüm adaylarına avantaj sağlamıştır. Sayısal 2 testinde de 3 maddenin sayısal bölüm adaylarına avantaj sağladığı tespit edilmiştir.

Başusta (2013) PISA 2006 testi maddeleri üzerinde farklı alt gruplarda kültüre ve dile göre Mantel Haenszel (MH), lojistik regresyon (LR), Alan indeksleri yöntemleriyle DMF analizleri uygulamıştır. Kanada, Avustralya, İngiltere ve Türkiye örneklemelerinin seçildiği araştırmada DMF'nin olası nedenlerini belirlemek için uzman görüşlerine başvurulmuştur. Elde edilen sonuçlara göre kullanılan DMF belirleme yöntemi değiştikçe DMF içeren madde sayısı farklılaşmaktadır. Dilsel ve kültürel farklılıkların artmasıyla da DMF içeren madde sayısının arttığı tespit edilmiştir.

Demir (2013) PISA 2009 Matematik okuryazarlığı alt testinde yer alan maddelerin değişen madde fonksiyonu (DMF) içerip içermediğini cinsiyet ve kültür değişkenlerine göre araştırmışlardır. DMF belirleme çalışmaları için Mantel Haenszel, Lojistik Regresyon ve SIBTEST yöntemleri kullanılmıştır. Cinsiyet değişkeni için veri seti, Türkiye'de 332 öğrencinin 5. kitapçıktaki matematik okuryazarlığı alt testinde yer alan maddelere verdikleri cevaplar, kültür değişkeni için Türkiye, Almanya, Finlandiya

ve ABD'deki öğrencilerin 5. kitapçıktaki matematik okuryazarlığı testine verdikleri cevaplardan oluşmaktadır. Analiz sonuçlarına göre cinsiyet değişkenine göre 4 maddenin erkekler lehine, bir maddenin de kızlar lehine avantajlı olduğu tespit edilmiştir. Kültür değişkenine göre ise Türkiye ile Almanya'dan katılan öğrencilerden elde edilen veri setlerinde 16, Türkiye ile Finlandiya'dan katılan öğrencilerden elde edilen veri setlerinde 14, Türkiye ile ABD'den katılan öğrencilerden elde edilen veri setlerinde ise 18 maddenin DMF içerdiği belirlenmiştir.

Svetina ve Rutkowski (2014) gerçekleştirdikleri simülasyon çalışması ile eşleştirme değişkeni olan toplam puanın DMF'li maddelerden arındırılmasının I. Tip hataya etkisine bakmışlardır. DMF analizler sonucunda DMF'li maddeler eşleştirme değişkeni olan toplam puana dâhil olduğunda I. tip hatanın düştüğü gözlemlenirken DMF'li maddeler toplam puandan arındırıldıktan sonra ise I. tip hatanın arttığı belirlenmiştir.

Erdem (2015) tarafından TEOG Ortak Sınavında yer alan maddelerin kitapçık türü değişkenine göre DMF gösterip göstermediğinin incelendiği araştırmada DMF analizleri için Mantel-Haenszel (MH), Lojistik Regresyon (LR) ve SIBTEST yöntemleri kullanılmıştır. Araştırmanın çalışma grubu, 2014-2015 öğretim yılı Sonbahar Dönemi'nde TEOG Ortak Sınavına giren öğrenciler arasından 12.000 öğrencinin seçkisiz yöntemle seçilmesiyle oluşturulmuştur. Kitapçık türüne göre DMF analizleri, kitapçıkların ikili gruplar halinde karşılaştırılmasıyla gerçekleştirilmiştir. Elde edilen sonuçlara göre, Din Kültürü ve Ahlak Bilgisi, T.C. İnkılap Tarihi ve Atatürkçülük ve İngilizce alt testlerinde DMF içeren madde sayısının fazla olduğu, Türkçe ve Fen ve Teknoloji alt testlerinde ise daha az maddenin DMF içerdiği belirlenmiştir. Matematik alt testinde yer alan maddelerde ise kitapçık türüne göre DMF içeren madde tespit edilmemiştir. Kullanılan MH, LR ve SIBTEST yöntemleri ile elde edilen sonuçlar karşılaştırıldığında, SIBTEST yöntemi ile 120 maddenin 28'inin, MH yöntemi ile 17'sinin, LR yöntemi ile de 8'inin DMF içerdiği belirlenmiştir. Başka bir ifadeyle bu yöntemlerin DMF içeren madde tespitinde madde sayısı bakımından birbiriyle uyum içerisinde olmadığı görülmüştür. Yöntemlerin uyumunun aynı maddelerin DMF içerip içermediğine göre karşılaştırılmasında ise en yüksek düzeyde uyumun MH ve SIBTEST

yöntemlerinde, en az düzeyde uyumun ise MH-LR yöntemlerinde olduğu tespit edilmiştir.

Lee ve Geisinger (2016) gerçekleştirdikleri simülasyon çalışmalarında, eşleştirme kriteri olan arındırmanın (purification) değişen madde fonksiyonu (DMF) gösteren maddelerin tespitine olan etkisini Mantel– Haenszel yöntemiyle araştırmışlardır. Testin uzunluğu, DMF gösteren maddelerin oranı, örneklem büyüklüğü, referans ve odak grubunun oranı ve iki grup arasında ortalama yetenek farkının varlığı manipüle edilen değişkenlerdir. DMF analizleri arındırma öncesi ve sonrası gerçekleştirilmiştir. Sonuçlara göre eşleştirme kriteri olan toplam puanın arındırılmasının genel olarak DMF saptanmasında gücün gelişmesine katkıda bulunduğunu göstermiş ve ayrıca gruplar arasında ortalama bir yetenek farkı olmadığında, Tip I hata oranlarında hafif bir düşüş gözlenmiştir.

Sünbül ve Ömür Sünbül (2016) simülasyon çalışmalarında Klasik Test Kuramı ve Madde Tepki Kuramı'na dayalı çeşitli DMF belirlenme yöntemlerin belirli koşullardaki I. tip hata ve güç oranlarını belirlemeyi amaçlamıştır. KTK'ya dayalı yöntemlerde, farklı örneklem büyüklüğünün, farklı madde sayısının ve referans grubundaki örneklem büyüklüğünün toplam örneklem büyüklüğüne oranının farklılaşarak kullanıldığı durumların 1. tip hata oranları ortalamalarını belirgin bir şekilde değiştirmedeği görülmüştür. MTK'ya dayalı yöntemlerde ise, örneklem büyüklüğü arttıkça I. Tip hata oranlarının ortalamasının azaldığı, madde sayısının arttıkça I. tip hata oranlarının ortalamasının değişmediği ve referans ve odak gruplardaki örneklem büyüklüğü oranlarının eşit olduğu koşulda da I. Tip hata oranları ortalamasının daha düşük olduğu gözlenmiştir. Güç oranlarının belirlenmesi için yapılan çalışmada, KTK ve MTK'ya dayalı yöntemlerde, örneklem büyüklüğünün ve DMF içeren madde sayısının artmasının güç oranlarının ortalamasının da arttığı belirlenmiştir. Tüm yöntemlerde farklı madde sayısının güç oranlarının ortalamalarını çok fazla değiştirmedeği, DMF içeren madde oranının arttırılmasının ise güç oranlarının ortalamasını azalttığı tespit edilmiştir.

Yıldırım (2017) araştırmasında, PISA 2009 okuma becerileri alanındaki maddelerin Türkiye, Fransa ve ABD örneklemi DMF'li madde içerip içermediğini

eşleştirme değişkeni olan toplam puan DMF'li maddelerden arındırılmadan önce ve arındırıldıktan sonra tek değişkenli ve çok değişkenli eşleştirme yöntemlerine dayalı olarak belirlemeyi ve madde yanlılığı kaynaklarını ortaya koymayı amaçlamıştır. Elde edilen sonuçlara göre tek değişkenli DMF analizlerinin SIBTEST ve lojistik regresyon yöntemleri arasında paralellik gösterdiği ancak kullanılan yöntemlere göre DMF düzeylerinin farklılaştığı belirlenmiştir. Ayrıca toplam puana ek olarak eşleştirme değişkenlerinin kullanımında değişken sayısı arttıkça I. tip hatanın azaldığı sonucuna ulaşılmıştır. Eşleştirme değişkeni olan toplam puan hesaplanırken DMF'li maddelerden arındırılmasının da tutarlı sonuçlar vermemekle birlikte DMF'li olarak etiketlenen madde sayısında ve maddelerin DMF düzeylerinde değişimlere yol açtığı gözlenmiştir.

III. BÖLÜM

3. Yöntem

Bu bölümde, araştırmanın modeli, simülasyon koşulları, verilerin üretilmesi ve verilerin analizine yer verilmiştir.

3.1. Araştırmanın Modeli

Bir testin DMF'li madde içeren ve DMF'li maddeden arındırılmış formlarından kestirilen yetenek kestirimlerinin farklı madde sayısı, farklı örneklem büyüklüğü ve farklı DMF oranı koşulları altında karşılaştırılmasının amaçlandığı bu araştırma ilişkisel tarama modelinde bir araştırmadır. Karasar (2016)'a göre ilişkisel tarama modeli, birden fazla değişken arasında birlikte değişimin varlığını ve/veya derecesini belirlemeyi amaçlayan araştırma modelidir.

3.2. Simülasyon Koşulları

Bu araştırmada çeşitli koşullar altında bir testin DMF'li madde içeren ve DMF'li maddeden arındırılmış formlarından kestirilen yetenek kestirimlerinin karşılaştırılması, simülasyon çalışmasıyla gerçekleştirilmiştir. Bazen istatistiksel karmaşıklıkları gidermek için simülasyon çalışmaları kolaylıklar sağlayabilmektedir. Simülasyon çalışmaları sayesinde bilinen ve istenen özelliklere uygun ve yeterli olarak belirtilen koşullar altında veri elde edilebilmektedir (Hallgren, 2013). Feinberg ve Rubrigh (2016) 'e göre gerçek parametreler pratikte bilinmediği için istenilen parametrelerin özelliklerine simülasyondan önce araştırmacılar tarafından karar verilir. Bu da araştırmacının istediği modeli özelleştirmesine fırsat sağlar. Bu yüzden bu çalışmaların dikkatle yürütülmesi gerekir. Simülasyon çalışmaları ile araştırmacılar pratikte bilinmeyen ancak kendilerinin

gerçek parametrelerini kestirdikleri parametreler ile karşılaştırma imkanı bulabilir. Bazı durumlarda gerçek veriler yerine simülasyon verilerin kullanılmasının avantajı araştırmacının düşündüğü varsayımların sağlanmış olmasıdır. Başka bir ifadeyle simülasyon çalışmalarıyla araştırmacı çalışmak istediği koşulları gerekli temellere dayandırarak kendisi belirleyebilmektedir.

Bu simülasyon çalışması için veri üretiminde sabit tutulan ve manipüle edilen koşullar aşağıda açıklanmıştır.

Sabit Koşullar:

Bu çalışmada simülasyon verileri iki kategorili yapıya göre puanlanan maddelere uygun olacak biçimde üretilmiştir. DMF bulunan maddelerin tek biçimli olması çalışmanın sabit tutulan diğer bir koşuludur. Ayrıca tüm koşullarda bireylerin yetenek parametreleri ortalaması 0 ve standart sapması 1 olan standart normal dağılıma uygun elde edilmiştir. MTK modeline uygun verilerin üretilmesi ise üç parametrelili lojistik modele dayalı olarak yapılmıştır. Bu model için a, b ve c parametrelerinin ortalaması ve standart sapması ya da minimum ve maksimum değerleri belirlenerek bu değerler arası veri üretimi gerçekleştirilmiştir. Gerçekleştirilen veri üretiminde Sünbül ve Ömür Sünbül (2016)'ün çalışmalarında ürettikleri verilerin parametrelerine benzer olarak a parametresi ortalaması 0.8 standart sapması 0.04 olan normal dağılımdan elde edilmiştir. Uniform dağılımdan elde edilen b parametresinin de minimum değeri -2, maksimum değeri +2 olarak belirlenmiştir. Son olarak c parametresinin değer aralığı da 0.2 ile 0.3 olarak tespit edilip veri üretilmiştir.

Manipüle Edilen Koşullar:

Alanyazında DMF üzerine yapılan araştırmalarda; madde sayısı başka bir ifadeyle test uzunluğu, örneklem büyüklüğü ve DMF içeren madde oranı gibi birçok değişkenin DMF üzerinde etkisi olduğu görülmektedir (Clauser, Mazor ve Hambleton, 1993; Narayanan ve Swaminathan 1996; French ve Maller 2007; Atar ve Kamata, 2011).

Bu çalışmada da bu koşullar manipüle edilerek veri üretimi ilgili koşulların belirlenen düzeylerine göre gerçekleştirilmiştir.

Madde sayısı (k): Bu çalışmada madde sayısı koşulu için $k=40$ ve $k=60$ olacak şekilde iki düzey belirlenmiştir. Standart başarı ve yetenek testleri genel olarak 35 madde ile 80 madde arasında değişkenlik göstermektedir (Narayanan ve Swaminathan 1996; French ve Maller 2007). Ulusal ve uluslar arası standart başarı ve yetenek testleri de göz önüne alındığında bu çalışmanın madde sayısı düzeyleri 40 ve 60 olarak kararlaştırılmıştır.

Örneklem büyüklüğü (n): Çalışmanın örneklem büyüklüğü $n=1000$ ve $n=2000$ olmak üzere iki düzey olarak belirlenmiştir. Örneklem büyüklüğünün 1000 olduğu koşulda referans ve odak gruplarının kişi sayısı 500'er; örneklem büyüklüğünün 2000 olduğu koşulda ise referans ve odak gruplarının kişi sayısı 1000'er olarak ele alınmıştır. MTK'ya dayalı DMF belirleme yöntemleri üzerine yapılan simülasyon çalışmalarında örneklem büyüklüğünün en az her grup için 200-250 kişi, toplamda ise en az 600 kişi olmasının yeterli olduğu görülmüştür (Narayanan ve Swaminathan 1996; French ve Maller 2007; Atar ve Kamata, 2011).

DMF'li madde oranı: Jodoin ve Gierl (2001)'e göre DMF'li madde oranının yüksek olması, başka bir ifadeyle test içindeki DMF'li madde sayısının fazla olması testin geçerliğine tehdit oluşturacağı için çalışmanın DMF'li madde oranı $d=5\%$ ve $d=10\%$ olmak üzere iki düzeyde ele alınmıştır. Ayrıca DMF üzerine yapılan çalışmalarda üzerinde araştırma yapılan testlerde birden çok sayıda DMF'li maddeye rastlanıldığı görülmektedir. Bu sebeplerle belirlenen DMF'li madde oranı 5% ve 10% düzeylerine göre, madde sayısının 40 olduğu koşulda DMF'li madde sayıları sırasıyla 2 ve 4; madde sayısının 60 olduğu koşulda DMF'li madde sayıları sırasıyla 3 ve 6 olarak belirlenmiştir.

3.3. Verilerin Üretilmesi

Bu simülasyon çalışmasında veri üretimi üç parametrelili lojistik modele dayalı olarak R 3.4.1 programına kod yazılmasıyla gerçekleştirilmiştir. Ele alınan her bir koşul için 50 tekrarlama yapılmıştır.

Yukarıda çalışmanın veri üretimi için sabit tutulan ve manipüle edilen koşulları açıklanmıştır. Manipüle edilen madde sayısı, örneklem büyüklüğü ve DMF'li madde oranı koşullarının her birinin düzeylerine göre gerçekleştirilen simülatif veri üretiminin çalışma planı Tablo 3.1 'de gösterilmiştir.

Tablo 3.1. Simülatif veri üretim planı.

K	Madde Sayısı	Örneklem Büyüklüğü	DMF'li Madde Oranı
1	40	1000 (R:500/O:500)	%5
2	60	1000 (R:500/O:500)	%5
3	40	2000 (R:1000/O:1000)	%5
4	60	2000 (R:1000/O:1000)	%5
5	40	1000 (R:500/O:500)	%10
6	60	1000 (R:500/O:500)	%10
7	40	2000 (R:1000/O:1000)	%10
8	60	2000 (R:1000/O:1000)	%10

K: Koşul

R: Referans grup

O: Odak grup

Tablo 3.1. incelendiğinde manipüle edilen her bir koşulun iki düzeyi için çaprazlanan olası tüm koşullara uygun veri üretiminin planlandığı görülmektedir.

Planlanan koşullara uygun veriler her koşulda a parametresi ortalaması 0.8 standart sapması 0.04 olan normal dağılımdan elde edilmiştir. Ayrıca b parametresinin minimum değeri -2, maksimum değeri +2 olarak belirlenmiştir. Son olarak c parametresinin değer aralığı da 0.2 ve 0.3 arası olarak tespit edilip veri üretilmiştir.

Manipüle edilen koşul olan DMF'li madde oranının belirlenen düzeylerine göre DMF'li madde üretilmesi için gerekli madde sayısına uygun olacak biçimde ilgili maddelerin b parametrelerine DMF miktarı olarak 0.75 eklenmiştir. Başka bir ifadeyle referans grubun b parametreleri üzerinde istenen DMF'li madde sayısı kadar parametreye 0.75 eklenerek odak gruba ait veriler üretilmiştir. Böylece bireylerin yetenek düzeyleri farklılaştırılmadan o maddelerin güçlük düzeyleri farklılaştırılarak DMF'li maddeler elde edilmiştir. Bu durum araştırmacının çalışması için belirlediği miktarlarda her bir koşula göre gerçekleştirilmiştir. Kısaca ilgili koşula göre üretilen veri setlerinde test içerisinde koşulun düzeyinin gerektirdiği miktarda madde DMF'li olarak üretilmiştir.

3.4. Verilerin Analizi

Bu simülasyon çalışması için veri üretiminin R 3.4.1 programıyla gerçekleştirildiği yukarıda açıklanmıştır. Çalışmanın amacı doğrultusunda ve yukarıda açıklanan koşullara uygun olarak DMF'li madde içeren veri setleri üretilmesi için R programı içerisinde “difR” paketi kullanılmıştır. Her bir koşula göre tekrarlı olarak üretilen veri setlerinde bireylerin test cevaplarına göre yetenek kestirimleri MTK'nın üç parametrelili lojistik modeline dayalı olarak R programının “lrm” paketi ile gerçekleştirilmiştir. Her bir koşulun aynı testinden DMF'li maddeler çıkarılarak bireylerin yetenekleri tekrar kestirilmiştir. Başka bir ifadeyle bir testin hem DMF'li madde içeren hem de DMF'li maddeden arındırılmış formlarından yetenek kestirimleri ayrı ayrı yapılmıştır. Testin DMF'li madde içeren formuna ait yetenek kestirimleri yani

bireylerin θ (theta) deęerleri θ_1 olarak ele alınırken testin DMF'li maddeden arındırılmış formunda bireylerin yetenek kestirimleri θ_2 olarak belirlenmiştir.

Her bir koşulun 50 tekrarı için θ_1 ve θ_2 deęerleri arasında ilişki olup olmadığına, ilişki varsa düzeyine bakmak için SPSS 22.0 programı kullanılmıştır. SPSS 22.0 programıyla her tekrar için θ_1 ve θ_2 arasındaki ilişkinin varlığı Spearman sıra farkları korelasyonu analizi gerçekleştirilerek incelenmiştir. Aynı koşul içinde her bir tekrarlama için Spearman sıra farkları korelasyonu analizi sonucunda elde edilen korelasyon katsayılarının (r_s) ortalaması hesaplanmak istenmiştir. Korelasyon katsayılarının ortalamasını hesaplamada daha net sonuçlar elde etmek için Corey, Dunlap ve Burke (1998)'nin önerdiği Fisher-Z dönüşümü uygulanmıştır. Bunun için öncelikle her bir r_s katsayısı ayrı ayrı Fisher-Z dönüşümü ile z deęerine çevrilmiş, ardından çevrilen bu z deęerlerinin ortalaması hesaplanmış ve bulunan ortalama z deęeri tekrar Fisher-Z dönüşümü ile r_s 'ye çevrilmiştir. Bu şekilde bir koşuldaki 50 tekrarlama ile elde edilen θ_1 ve θ_2 deęerlerinin ilişkisi ortalama bir korelasyon katsayısı bulunarak gözlenmiştir. İlişki düzeyine bakmada bu işlem 8 ayrı koşul için gerçekleştirilmiştir. Başka bir ifadeyle 8 ayrı koşulun 50 tekrarlama ile elde edilen korelasyon katsayıları Fisher-Z dönüşümü kullanılarak ortalama bir korelasyon katsayısına dönüştürülmüştür. Araştırmanın bulgular bölümünde verilen r_s deęerleri bu işlemler sonucunda elde edilen ortalama spearman korelasyon katsayısını göstermektedir.

IV. BÖLÜM

4. Bulgular ve Tartışma

Bu bölümde araştırmacı tarafından belirlenen sabit ve manipüle edilen koşullara uygun üretilen verilerin analizi sonucunda elde edilen bulgular yer almaktadır.

Toplam sekiz koşul için testin DMF'li madde içeren formuna ait yetenek kestirimleri (θ_1) ile DMF'li maddeden arındırılmış formuna ait yetenek kestirimleri (θ_2) arasındaki ilişki düzeyinin incelendiği, Spearman Sıra Farkları Korelasyon Analizi ile elde edilen bulgular öncelikle genel olarak yorumlanmış ardından ayrı ayrı verilen bulgular manipüle edilen madde sayısı, örneklem büyüklüğü ve DMF oranı koşullarına göre alt problemlere uygun olarak yorumlanmıştır. Başka bir ifadeyle testin DMF'li madde içeren ve DMF'li maddeden arındırılmış formlarından kestirilen yetenek kestirimlerinin (θ_1 ve θ_2) araştırmanın tüm koşullarındaki ilişki düzeyi Spearman Sıra Farkları Korelasyon Katsayısı (r_s) kullanılarak önce genel sonra da alt problemlere göre ayrı ayrı yorumlanmıştır.

Aşağıdaki tabloda tüm koşullar için θ_1 ve θ_2 değişkenleri arasındaki korelasyon katsayı değerleri birlikte verilmiştir.

Tablo 4.1. Tüm koşullardan elde edilen korelasyon katsayıları

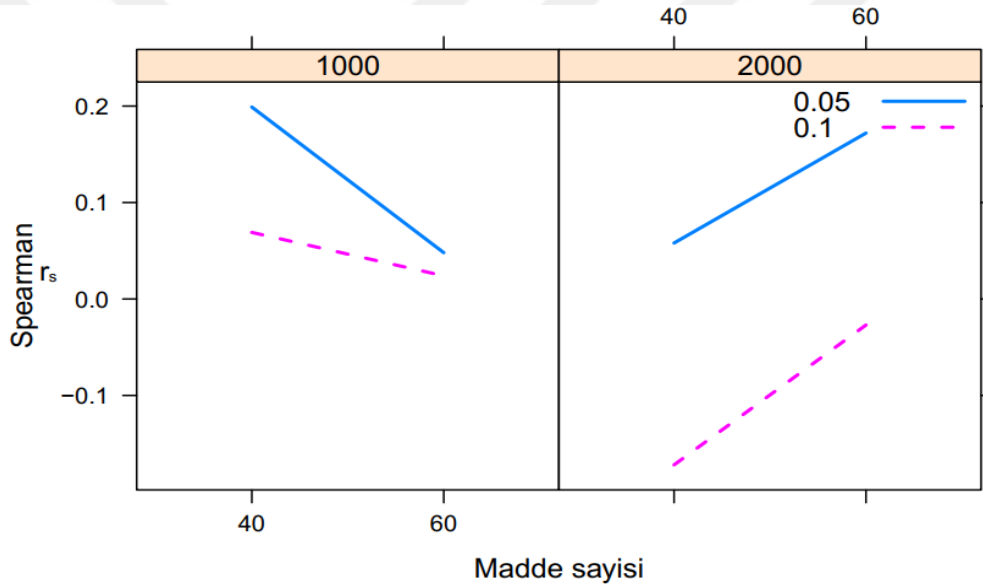
K	Örneklem Büyüküğü	DMF Oranı	Madde Sayısı	r_s
1	1000	%5	40	0.199
2	1000	%10	40	0.069
3	2000	%5	40	0.058
4	2000	%10	40	-0.143
5	1000	%5	60	0.048
6	1000	%10	60	0.024
7	2000	%5	60	0.172
8	2000	%10	60	-0.027

Not 1: K her çaprazlama için ilgili koşulu ifade etmektedir.

Not 2: r_s spearman sıra farkları korelasyon değerini belirtmektedir ($p<0.01$).

Tablo 4.1. incelendiğinde sekiz koşul için korelasyon katsayıları (r_s) tablodaki sıralamaya göre 0.199; 0.069; 0.058; -0.143; 0.048; 0.024; 0.172 ve -0.027 olarak bulunmuştur. Bu sekiz bulgunun tamamının 0'a yakın değerler olduğu görülmektedir. Bu bulgular da ilgili koşulların her biri için θ_1 ve θ_2 arasında ilişkinin olmadığını ya da yok denecek kadar az ilişkinin olduğunu ifade etmektedir.

Çaprazlanan tüm koşullara göre ilişkinin düzeyini tablo 4.1 dışında, daha detaylı ve net biçimde tüm koşulların birbiriyle etkileşimli olarak verildiği şekil üzerinde incelemenin faydalı olacağı düşünülerek şekil 4.1 aşağıda gösterilmiştir.



Şekil 4.1. Tüm koşulların korelasyon katsayıları

Şekil 4.1. incelendiğinde dikey ekseninde spearman korelasyon katsayı değerleri (r_s) yer alırken yatay ekseninde ise madde sayısı 40 ve 60 olmak üzere iki düzeyde görülmektedir. Ayrıca yatay ekseninde örneklem büyüklüğü ise 1000 ve 2000 düzeylerinde yer almaktadır. Grafikler ise DMF oranının %5 = 0.05 ve %10 = 0.1 olarak belirlenen düzeylerine göre ayrı ayrı çizilmiştir. Şekil 4.1. genel olarak incelendiğinde yukarıda tablo 4.1. üzerinde verilen tüm r_s değerlerinin araştırmacının koşullarının düzeylerine göre yaklaşık olarak 0.2 ile -0.15 arasında değiştiği görülmektedir. Başka bir ifadeyle araştırmacının tüm çalışma koşullarına göre dikey ekseninde yer alan ilişki düzeyinin maksimum ve minimum olduğu koşullar da dahil olmak üzere her birinin pozitif yönlü

0.3' ten küçük ya da negatif yönlü -0.3' ten büyük olduğu bulunmuştur. Bu bulgular da belirlenen tüm koşullardan hangisi olursa olsun ayrı ayrı kestirilen yetenek kestirimleri arasında genel bir ilişkisizliğin varlığını göstermektedir. Yani aynı testin DMF'li madde içeren formuna göre oluşan bireylerin sıralamaları ile DMF'li maddeden arındırılmış formuna göre oluşan sıralamalar arasında ilişki yoktur. Daha açık bir ifadeyle testte bulunan DMF'li maddeler testten çıkarılıp yeniden kestirimler yapıldığında testi alan bireylerin sıralamaları değişmektedir.

Bir testte DMF'li madde ya da maddelerin bulunmasının o testin geçerliğine tehdit oluşturduğu üzerine daha önce detaylı açıklamalar yapılmıştı. Böyle bir durumda testte DMF'li madde varken geçerlik bu durumdan etkileneceği için o testten elde edilen puanların yerinde ve isabetli olması tartışmalı hâle gelmektedir. Yani bir testte DMF'li madde bulunması istenen bir durum değildir. Öyle ki DMF'li madde varken de bu maddelerin çıkarılarak yeniden puanlama yapılmasıyla testi alan bireylerin sıralamaları değişiyorsa bu durum gerçek puana yakın bir puan elde edilemediğini göstermektedir. Gerçek puana yakın bir puan elde edilememesi de gerek ulusal gerekse uluslararası birçok alanda, sıralama ya da seçme amaçlı uygulanan sınavlardan alınan hayati kararların sorgulanması gerektiğini ortaya çıkarabilmektedir.

Bakan Kalaycıoğlu ve Kelecioğlu (2011), Başusta (2013), Çepni (2011), Demir (2013), Doğan ve Öğretmen (2008), Erdem (2015), Gök, Kelecioğlu ve Doğan (2010), Öğretmen (2006), Yıldırım (2017) yaptıkları çalışmalarda SBS, ALES, PISA, TEOG, ÖSS gibi ulusal ya da uluslararası başarı testlerinde DMF'li maddelerin yer aldığı sonucuna ulaşmışlardır. Bu testlerde DMF'li maddelerin bulunması testlerin geçerliği için tehdit oluşturabilmektedir. Başka bir ifadeyle testte DMF'li madde bulunması o testin ölçmek istediği özelliğe sahip olanla olmayanı ayırt etmekte sorun yaşayacağını göstermektedir. Bu çalışma ile de DMF'li maddelerin testten arındırılmasıyla yapılan kestirimler sonucunda bireylerin başarı sıralamalarında farklılıklar olduğu bulgusuna ulaşılmıştır. Dolayısıyla alanyazında DMF'li maddelerin tespit edildiği testlerin DMF'li maddeden arındırılmasıyla yeniden analizler yapıldığı takdirde bireylerin başarı sıralamalarında değişiklikler meydana gelebileceği düşünülmektedir. Bu durum da testteki DMF'li maddelerin testten arındırılmasının önemli bir etkisini göstermektedir. Bu

anlamda bu çalışmanın bulguları aynı zamanda DMF çalışmaları için testlerin DMF’li maddelerden arındırılması gerektiği önerilerine (French ve Maller, 2007; Holland ve Thayer, 1988, Lee ve Geisinger, 2016, Zumbo 1999) paralellik göstermektedir.

4.1. Birinci Alt Probleme İlişkin Bulgular ve Yorumlar

Aynı örneklem büyüklüğü ve DMF oranı ile farklı madde sayısı (k=40 ve k=60) altında bir testin DMF’li madde içeren ve DMF’li maddeden arındırılmış formlarından kestirilen yetenek kestirimleri ilişkisi ne düzeydedir?

Yukarıda verilen tablo 4.1.’deki aynı spearman sıra farkları korelasyon katsayısı (r_s) değerleri kullanılarak tablo birinci alt probleme uygun olacak biçimde yeniden düzenlenip tablo 4.2. oluşturulmuştur. Aşağıda verilen tablo 4.2. ile örneklem büyüklüğü ve DMF oranı aynıyken madde sayısının farklılaşmasıyla korelasyon katsayılarının nasıl değiştiği daha anlaşılır biçimde görülmektedir.

Tablo 4.2. Madde sayısının farklılaştırılmasıyla elde edilen korelasyon katsayıları

K.	Örneklem Büyüklüğü	DMF Oranı	Madde Sayısı	r_s
1	1000	%5	40	0.199
2	1000	%5	60	0.048
3	2000	%5	40	0.058
4	2000	%5	60	0.172
5	1000	%10	40	0.069
6	1000	%10	60	0.024
7	2000	%10	40	-0.143
8	2000	%10	60	-0.027

Not 1: K her çaprazlama için ilgili koşulu ifade etmektedir.

Not 2: r_s spearman sıra farkları korelasyon değerini belirtmektedir ($p<0.01$).

Not 3: Bulgular aynı örneklem büyüklüğü ve DMF oranı ile farklı madde sayısı durumlarına göre ikili olarak gruplandırılmıştır.

Tablo 4.2. incelendiğinde, 1. ile 2. koşulların 1000 örneklem büyüklüğü ve %5 DMF oranı altında, 3. ile 4. koşulların 2000 örneklem büyüklüğü ve %5 DMF oranı altında, 5. ile 6. koşulların 1000 örneklem büyüklüğü ve %10 DMF oranı altında, 7. ile 8. koşulların 2000 örneklem büyüklüğü ve %10 DMF oranı altında madde sayılarının 40 ve

60 olacak biçimde farklılaştırılarak hesaplanan spearman sıra farkları korelasyon katsayı (r_s) değerlerinin verildiği görülmektedir.

1. ve 2. koşullarda madde sayısı 40 iken r_s değeri 0.199; madde sayısı 60 iken r_s değeri 0.048 bulunmuştur. Bu bulgu madde sayısı 40 olduğunda ilişki düzeyi yok denecek kadar az iken; madde sayısı arttıkça ilişki düzeyinin 0'a daha yakın olacak biçimde düştüğünü ve ilişkisizliğin devam ettiğini göstermektedir. Sonuç olarak örneklem büyüklüğünün 1000, DMF oranının %5 olduğu koşulda madde sayısının artırılmasının yetenek kestirimleri arasındaki ilişki durumuna manidar bir etkisi olmamıştır.

3. ve 4. koşullarda madde sayısı 40 iken r_s değeri 0.058; madde sayısı 60 iken r_s değeri 0.172 bulunmuştur. Bu bulguya göre madde sayısı 40 olduğunda yetenek kestirimleri arasında ilişki yok iken; madde sayısı arttıkça ilişki düzeyinde artış gözlemlense bile bu ilişki düzeyi hâlâ değişkenler arasında bir ilişkinin varlığını gösterecek değerde değildir. Sonuç olarak örneklem büyüklüğünün 2000, DMF oranının %5 olduğu durumda madde sayısının artmasıyla ilişki durumunda bir farklılık görülmemiştir.

5. ve 6. koşullarda madde sayısı 40 iken r_s değeri 0.069; madde sayısı 60 iken r_s değeri 0.024 bulunmuştur. Bu bulgu madde sayısı 40 olduğunda yetenek kestirimleri arasında ilişki olmadığını; madde sayısı arttıkça ilişki düzeyinin 0'a daha yaklaştığını, yani ilişkisizliğin devam ettiğini göstermektedir. Sonuç olarak örneklem büyüklüğünün 2000, DMF oranının %10 olduğu durumda madde sayısının artmasıyla ilişki düzeyinde bir farklılık görülmemiştir.

7. ve 8. koşullarda madde sayısı 40 iken r_s değeri -0.143; madde sayısı 60 iken r_s değeri -0.027 bulunmuştur. Bu bulguya göre madde sayısı 40 olduğunda ilişkinin düzeyi yok denecek kadar az iken; madde sayısı arttıkça ilişki düzeyinin artarak 0'a yaklaştığı ve yine ilişkisizliğin devam ettiği görülmektedir. Sonuç olarak örneklem büyüklüğünün 2000, DMF oranının %10 olduğu durumda madde sayısının artmasıyla ilişki düzeyinde bir farklılık görülmemiştir.

Şekil 4.1. madde sayısının farklılaşma durumuna göre tekrar incelendiğinde, 1000 örneklem büyüklüğünde DMF oranının %5 ve %10 olduğu iki farklı grafikte de madde sayısının 40'tan 60'a artırıldığında spearman sıra farkları korelasyon katsayısı (r_s) değerinin azalarak 0'a yaklaştığı görülmektedir. Madde sayısı arttıkça yetenek kestirimleri arasındaki ilişki düzeyinin azalarak 0'a yaklaşması, bireylerin testin DMF'li madde ve DMF'li maddeden arındırılmış olmak üzere iki farklı formundaki sıralamaları arasında bir ilişki bulunmadığını; dolayısıyla bu sıralamaların daha fazla değiştiğini göstermektedir. 2000 örneklem büyüklüğünde ise DMF oranı %5 ve %10 iken madde sayısının 40'tan 60'a artırılmasıyla yetenek kestirimleri arasındaki ilişki düzeyi artış gösterse bile bu artış yine de herhangi bir ilişkinin varlığından söz edebilmek için yeterli değildir. Çünkü hem 40 hem de 60 madde sayısında korelasyon katsayılarının ilişkinin varlığından söz edebilecek biçimde farklılaşmaması söz konusudur. Kısaca, koşullar kendi aralarında ayrı ayrı incelendiğinde bu koşullarda örneklem büyüklüğü ve DMF oranı aynı kalırken madde sayısı artsa bile ilişkinin düzeyinde manidar bir farklılık görülmemektedir. Başka bir ifadeyle, aynı örneklem büyüklüğü ve DMF oranı ile farklı madde sayısı ($k=40$ ve $k=60$) altında bir testin DMF'li madde içeren ve DMF'li maddeden arındırılmış formlarından kestirilen yetenek kestirimleri arasında manidar bir ilişki bulunmamaktadır. Manidar bir ilişki olmaması da testin DMF'li maddelerden arındırılmasının bireylerin testteki sıralamalarını değiştirdiğini göstermektedir.

4.2. İkinci Alt Probleme İlişkin Bulgular ve Yorumlar

Aynı madde sayısı ve DMF oranı ile farklı örneklem büyüklüğü ($n=1000$ ve $n=2000$) altında bir testin DMF'li madde içeren ve DMF'li maddeden arındırılmış formlarından kestirilen yetenek kestirimleri ilişkisi ne düzeydedir?

Araştırmanın spearman sıra farkları korelasyon katsayısı (r_s) değerlerinin yer aldığı tablo 4.1.'deki aynı r_s değerleri kullanılarak tablo ikinci alt probleme uygun olacak biçimde yeniden düzenlenip tablo 4.3. oluşturulmuştur. Aşağıda verilen tablo 4.3. ile madde sayısı ve DMF oranı aynıyken örneklem büyüklüğünün farklılaşmasıyla korelasyon katsayılarının nasıl değiştiği daha anlaşılır biçimde görülmektedir.

Tablo 4.3. Örneklem büyüklüğünün farklılaştırılmasıyla elde edilen korelasyon katsayıları

K.	Örneklem Büyüklüğü	DMF Oranı	Madde Sayısı	r_s
1	1000	%5	40	0.199
2	2000	%5	40	0.058
3	1000	%5	60	0.048
4	2000	%5	60	0.172
5	1000	%10	40	0.069
6	2000	%10	40	-0.143
7	1000	%10	60	0.024
8	2000	%10	60	-0.027

Not 1: K her çaprazlama için ilgili koşulu ifade etmektedir.

Not 2: r_s , spearman sıra farkları korelasyon değerini belirtmektedir ($p < 0.01$).

Not 3: Bulgular aynı madde sayısı ve DMF oranı ile farklı örneklem büyüklüğü durumlarına göre ikili olarak gruplandırılmıştır.

Tablo 4.3. incelendiğinde, 1. ile 2. koşulların 40 madde sayısı ve %5 DMF oranı altında, 3. ile 4. koşulların 60 madde sayısı ve %5 DMF oranı altında, 5. ile 6. koşulların 40 madde sayısı ve %10 DMF oranı altında, 7. ile 8. koşulların 60 örneklem büyüklüğü ve %10 DMF oranı altında örneklem büyüklüğünün 1000 ve 2000 olacak şekilde farklılaştırılarak hesaplanan spearman sıra farkları korelasyon katsayı (r_s) değerlerinin verildiği görülmektedir.

1. ve 2. koşullarda örneklem büyüklüğü 1000 iken r_s değeri 0.199; örneklem büyüklüğü 2000 iken r_s değeri 0.058 bulunmuştur. Yani örneklem büyüklüğü 1000 olduğunda ilişki düzeyi yok denecek kadar az iken; örneklem büyüklüğü arttıkça ilişki düzeyi daha da düşerek 0'a çok yakın bir değer almıştır ve ilişkisizlik devam etmiştir. Sonuç olarak madde sayısının 40, DMF oranının %5 olduğu durumda örneklem büyüklüğünün artması yetenek kestirimleri arasındaki ilişki durumunu etkilememiştir.

3. ve 4. koşullarda örneklem büyüklüğü 1000 iken r_s değeri 0.048; örneklem büyüklüğü 2000 iken r_s değeri 0.172 bulunmuştur. Yani örneklem büyüklüğü 1000 olduğunda yetenek kestirimleri arasında ilişki yok iken; örneklem büyüklüğü arttıkça ilişki düzeyi artsa bile genel bir ilişkisizlik durumu bulunmaya devam etmektedir. Sonuç olarak madde sayısının 60, DMF oranının %5 olduğu durumda örneklem büyüklüğünün

artırılmasının yetenek kestirimleri arasındaki ilişki durumuna manidar bir etkisi olmamıştır.

5. ve 6. koşullarda örneklem büyüklüğü 1000 iken r_s değeri 0.069; örneklem büyüklüğü 2000 iken r_s değeri -0.143 bulunmuştur. Bu bulgu örneklem büyüklüğü 1000 olduğunda değişkenler arasında ilişki olmadığını, örneklem büyüklüğü arttıkça ilişki düzeyinin de düştüğünü ve ilişkisizliğin devam ettiğini göstermektedir. Sonuç olarak madde sayısının 40, DMF oranının %10 olduğu durumda örneklem büyüklüğünün artmasıyla ilişki düzeyinde bir farklılık görülmemiştir.

7. ve 8. koşullarda örneklem büyüklüğü 1000 iken r_s değeri 0.024; örneklem büyüklüğü 2000 iken r_s değeri -0.027 bulunmuştur. Yani örneklem büyüklüğü 1000 olduğunda ilişki düzeyi yok denecek kadar az iken; örneklem büyüklüğü arttıkça ilişki düzeyinde bir düşüş olduğu ancak ilişkisizliğin sürdüğü anlaşılmaktadır. Sonuç olarak madde sayısının 60, DMF oranının %10 olduğu durumda madde sayısının artmasıyla ilişki düzeyinde bir farklılık görülmemiştir.

Şekil 4.1. örneklem büyüklüğünün farklılaşma durumuna göre tekrar incelendiğinde, madde sayısının 40 olarak belirlendiği durumda örneklem büyüklüğü 1000'den 2000'e artırıldığında DMF oranının %5 ve %10 olduğu iki farklı grafikte de spearman sıra farkları korelasyon katsayısı (r_s) değerinde bir miktar azalma olduğu ancak bu azalmanın herhangi bir ilişkiden söz edebilecek düzeyde olmadığı görülmektedir. Bu durum madde sayısının 40, DMF oranının %5 ve madde sayısının 40 DMF oranının %10 olduğu koşullarda örneklem büyüklüğünün 1000'den 2000'e artırılmasının DMF'li madde içeren ve DMF'li maddeden arındırılmış test formları arasındaki ilişkisizliğe manidar bir etki oluşturmadığını göstermektedir. Dolayısıyla örneklem büyüklüğünün 1000 ya da 2000 olduğu koşullarda DMF'li maddelerin testten arındırılmasının bireylerin sıralamalarında değişiklik meydana getirdiği anlaşılmaktadır. Madde sayısının 60 olarak belirlendiği durumda örneklem büyüklüğü 1000'den 2000'e artırıldığında DMF oranının %5 olduğu grafikte spearman sıra farkları korelasyon katsayısı (r_s) değerinde bir miktar artış olduğu ancak bu artışın herhangi bir ilişkiden söz edebilecek düzeyde olmadığı görülmektedir. Bu koşulda DMF oranının %10 olduğu grafikte ise örneklem

büyükliğünün artırılmasının korelasyon katsayısında azalma oluşturduğu ancak bu değişikliğin ilişkisizliği ortadan kaldıracak düzeyde olmadığı fark edilmektedir.

Sonuç olarak koşullar kendi aralarında ayrı ayrı incelendiğinde madde sayısı ve DMF oranı aynı kalırken örneklem büyüklüğü artsa bile ilişkinin düzeyinde manidar bir farklılık görülmemiştir. Başka bir ifadeyle, aynı madde sayısı ve DMF oranı ile farklı örneklem büyüklüğü ($n=1000$ ve $n=2000$) altında bir testin DMF'li madde içeren ve DMF'li maddeden arındırılmış formlarından kestirilen yetenek kestirimleri arasında manidar bir ilişki bulunmamaktadır. Manidar bir ilişki olmaması da testin DMF'li maddelerden arındırılmasının bireylerin testteki sıralamalarını değiştirdiğini göstermektedir.

4.3. Üçüncü Alt Probleme İlişkin Bulgular ve Yorumlar

Aynı madde sayısı ve örneklem büyüklüğü ile farklı DMF oranı ($d=5\%$ ve $d=10\%$) etkisi altında bir testin DMF'li madde içeren ve DMF'li maddeden arındırılmış formlarından kestirilen yetenek kestirimleri ilişkisi ne düzeydedir?

Araştırmanın spearman sıra farkları korelasyon katsayısı (r_s) değerlerinin yer aldığı tablo 4.1.'deki aynı r_s değerleri kullanılarak tablo ikinci alt probleme uygun olacak biçimde yeniden düzenlenip tablo 4.4. oluşturulmuştur. Aşağıda verilen tablo 4.4. ile madde sayısı ve örneklem büyüklüğü aynıyken DMF oranının farklılaşmasıyla korelasyon katsayılarının nasıl değiştiği daha anlaşılır biçimde görülmektedir.

Tablo 4.4. DMF oranının farklılaştırılmasıyla elde edilen korelasyon katsayıları

K.	Örneklem Büyüklüğü	DMF Oranı	Madde Sayısı	r_s
1	1000	%5	40	0.199
2	1000	%10	40	0.069
3	1000	%5	60	0.048
4	1000	%10	60	0.024
5	2000	%5	40	0.058
6	2000	%10	40	-0.143
7	2000	%5	60	0.172
8	2000	%10	60	-0.027

Not 1: K her çaprazlama için ilgili koşulu ifade etmektedir.

Not 2: r_s spearman sıra farkları korelasyon değerini belirtmektedir ($p<0.01$).

Not 3: Bulgular aynı madde sayısı ve örneklem büyüklüğü ile farklı DMF oranı durumlarına göre ikili olarak gruplandırılmıştır.

Tablo 4.4. incelendiğinde, 1. ile 2. koşulların 1000 örneklem büyüklüğü ve 40 madde sayısı altında, 3. ile 4. koşulların 1000 örneklem büyüklüğü ve 60 madde sayısı altında, 5. ile 6. koşulların 2000 örneklem büyüklüğü ve 40 madde sayısı altında, 7. ile 8. koşulların 2000 örneklem büyüklüğü ve 60 madde sayısı altında DMF oranının %5 ve %10 olacak biçimde farklılaştırılarak hesaplanan spearman sıra farkları korelasyon katsayı (r_s) değerlerinin verildiği görülmektedir.

1. ve 2. koşullarda DMF oranı %5 iken r_s değeri 0.199; DMF oranı %10 iken r_s değeri 0.069 bulunmuştur. Bu bulgu DMF oranı %5 olduğunda ilişki düzeyinin çok düşük olduğunu; DMF oranı arttıkça ilişki düzeyinin 0'a daha yakın olacak biçimde düştüğünü ve ilişkisizliğin devam ettiğini göstermektedir. Sonuç olarak örneklem büyüklüğünün 1000, madde sayısının 40 olduğu koşulda DMF oranının artırılmasının yetenek kestirimleri arasındaki ilişki durumuna manidar bir etkisi olmamıştır.

3. ve 4. koşullarda DMF oranı %5 iken r_s değeri 0.048; DMF oranı %10 iken r_s değeri 0.024 bulunmuştur. Bu bulguya göre DMF oranı %5 olduğunda yetenek kestirimleri arasında ilişki yok iken; DMF oranı arttıkça ilişki düzeyinden bir değişiklik gözlemlenmemiş ve hâlâ değişkenler arasında bir ilişkinin varlığını gösterecek korelasyon katsayısı tespit edilememiştir. Sonuç olarak örneklem büyüklüğünün 1000,

madde sayısının 60 olduğu durumda DMF oranının artmasıyla değişkenler arasındaki ilişki durumunda bir farklılık görülmemiştir.

5. ve 6. koşullarda DMF oranı %5 iken r_s değeri 0.058; DMF oranı %10 iken r_s değeri -0.143 bulunmuştur. Bu bulgu DMF oranı %5 olduğunda yetenek kestirimleri arasında ilişki olmadığını; DMF oranı arttıkça ilişki düzeyinin daha da düştüğünü, ancak ilişkisizliğin devam ettiğini göstermektedir. Sonuç olarak örneklem büyüklüğünün 2000, madde sayısının 40 olduğu durumda DMF oranının artmasıyla değişkenler arasındaki ilişki düzeyinde bir farklılık görülmemiştir.

7. ve 8. koşullarda DMF oranı %5 iken r_s değeri 0.172; DMF oranı %10 iken r_s değeri -0.027 bulunmuştur. Bu bulguya göre DMF oranı %5 olduğunda ilişkinin düzeyi yok denecek kadar az iken; DMF oranı arttıkça ilişki düzeyinin 0'a daha yakın bir değerde olduğu ve ilişkisizliğin devam ettiği görülmektedir. Sonuç olarak örneklem büyüklüğünün 2000, madde sayısının 60 olduğu durumda DMF oranının artmasıyla ilişki düzeyinde bir farklılık görülmemiştir.

Şekil 4.1. DMF oranının farklılaşma durumuna göre tekrar incelendiğinde tüm koşullarda DMF oranının %5'ten %10'a çıkarılmasıyla aslında var olan ilişkisizliğin daha da düştüğü görülmektedir. Başka bir ifadeyle DMF oranı artırıldıkça spearman sıra farkları korelasyon katsayısı (r_s) 0'a daha yakın bir değerdedir. Bunun sebebinin ise DMF oranının artması nedeniyle DMF'li olarak işaretlenen madde sayısının artması ve DMF'li madde sayısının fazlalığının da değişkenler arasında daha az ilişkiye yol açması olduğu düşünülmektedir.

Sonuç olarak ilişki düzeyi bakımından tüm koşullarda DMF oranının %5'ten %10'a artırılmasının DMF'li madde içeren ve DMF'li maddeden arındırılmış test formları arasındaki ilişkisizliğe manidar bir etki oluşturmadığını göstermektedir. Dolayısıyla DMF oranının %5 ve %10 olduğu koşullarda DMF'li maddelerin testten arındırılmasının bireylerin sıralamalarında değişiklik meydana getirdiği anlaşılmaktadır. Madde sayısının 60 olarak belirlendiği durumda örneklem büyüklüğü 1000'den 2000'e artırıldığında DMF oranının %5 olduğu grafikte spearman sıra farkları korelasyon

katsayısı (r_s) deęerinde bir miktar artış olduęu ancak bu artışın herhangi bir iliřkiden söz edebilecek düzeyde olmadığı grlmektedir. Bařka bir ifadeyle, aynı rneklem byklę ve madde sayısı ile farklı DMF oranı ($d=5\%$ ve $d=10\%$) altında bir testin DMF'li madde ieren ve DMF'li maddeden arındırılmıř formlarından kestirilen yetenek kestirimleri arasında manidar bir iliřki bulunmamaktadır. Manidar bir iliřki olmaması da testin DMF'li maddelerden arındırılmasının bireylerin testteki sıralamalarını deęiřtirdięini gstermektedir.



V.BÖLÜM

5. Sonuç ve Öneriler

Bu bölümde araştırmadan elde edilen sonuçlara ve ileride yapılacak araştırmalara dair önerilere yer verilmiştir.

5.1. Sonuçlar

Bir testin DMF'li madde içeren ve DMF'li maddeden arındırılmış formlarından kestirilen yetenek kestirimlerinin farklı madde sayısı, farklı örneklem büyüklüğü ve farklı DMF oranı koşulları altında karşılaştırılmasının amaçlandığı bu simülasyon çalışmasında, R 3.4.1 programıyla çeşitli koşullar altında üretilen veriler kullanılarak gerçekleştirilen analizlerin sonuçları aşağıda verilmiştir.

Çeşitli düzeylerde DMF'li maddeler içerecek şekilde verilerinin üretildiği bir testin öncelikle DMF'li maddeler içeriyorken yetenek kestirimleri gerçekleştirilmiştir. Testin DMF'li maddeler içeren hâliyle kestirilen yetenek kestirimlerine θ_1 ismi verilerek veriler saklı tutulmuştur. Ardından bu testte yer alan DMF'li maddeler testten arındırılarak aynı şekilde yetenekler kestirilmiştir. Testin DMF'li maddeler içermeyen hâliyle kestirilen yetenek kestirimleri ise θ_2 şeklinde saklanmıştır. Son olarak da aynı testin θ_1 ve θ_2 adıyla elde edilmiş olan bu kestirimleri arasındaki ilişkiye bakılmıştır. Bu yetenek kestirimleri ilişkisine göre bireylerin sıralamalarının farklılaşp farklılaşmadığını tespit etmek amaçlandığı için spearman sıra farkları korelasyon analizi uygulanmıştır. Bu korelasyon analizi sonucu elde edilen katsayıların 0'a yakın olmasından dolayı yetenek kestirimleri arasında pozitif ya da negatif yönlü bir ilişki görülmemiştir. Yetenek kestirimleri arasında ilişki görülmemesi ise bireylerin test sonuçlarındaki sıralamalarının değiştiğini işaret etmektedir. Başka bir ifadeyle test DMF'li maddelerden arındırıldıktan sonra bireylerin testteki sıralamaları, bir önceki DMF'li madde içeren test formu

sıralamalarına göre farklılaşmıştır. Bu tespit, çeşitli koşulların araştırıldığı tüm alt problemlerde benzer şekilde olmuştur. Başka bir ifadeyle madde sayısının 40 ve 60, örneklem büyüklüğünün 1000 ve 2000, DMF oranının %5 ve %10 olarak çaprazlandığı 8 farklı koşulda da testin DMF'li maddelerden arındırılmasının bireylerin sıralamalarını değiştirdiğini belirlenmiştir.

Bir testin DMF'li maddelerden arındırılmasıyla testi alan bireylerin sıralamalarının farklılaşması o testin geçerliğini yani özelliğe sahip olanla olmayana ayırt etme derecesini problemleri hâle getirebilecektir. Öyle ki testte DMF'li madde bulunması testin geçerliğine önemli bir tehdit oluştururken bu maddelerin testten çıkarılmasıyla bireylerin sıralamaları değişiyorsa, yapılan arındırma işleminin önemli bir etkisinin olduğu görülmektedir. Bu durum, gerek ulusal gerekse de uluslararası düzeyde bireyler hakkında hayati kararların alındığı, sonuçlarına bakılarak seçme ve yerleştirme işlemlerinin gerçekleştirildiği sınavların bireyler arasındaki farklılıkları ölçme derecelerinin sorgulanabilir olabileceğini gösterebilmektedir.

Sonuç olarak bir sınavda kullanılan ölçme aracının testi alan herhangi bir gruba avantaj sağlamaması gerekir. Bazı durumlarda bireylerin ölçmek istediğimiz özelliklerine başka değişkenler karışabilir. Bu değişkenler arasında cinsiyet, okul türü, sosyo-ekonomik düzey, etnik köken vb. yer alır (Atalay Kabasakal, 2014). Testin ölçmek istediği yapı ile ilişkisiz değişkenlerin test puanlarına etkisi geçerlik üzerinde bir tehdittir ve test puanlarının yanlı olmasına yol açar (Camilli ve Shephard, 1994). Yanlılığın belirlenmesinde ise koşulan ilk adım bu amaçla geliştirilmiş ve fazla sayıda yöntemi bulunan DMF analizleridir.

Herhangi bir testteki maddelerin kısmen dahi DMF içermesi durumunda istenmeyen sonuçlar elde edilebilir. Bu istenmeyen sonuçlardan bir tanesi de DMF'nin madde parametresi kestirimini doğrudan etkilemesidir (Han, 2008). Bir diğer istenmeyen sonuç ise yetenek parametrelerinin hatalı kestirimidir (Atalay Kabasakal, 2014). Madde ve yetenek parametrelerinin hatalı kestirimi sonucu bu parametrelere dayalı yapılan birçok istatistiksel çalışmaların sonuçlarından kuşku duyulması da kaçınılmazdır. Alanyazında DMF'li maddelerin testte bulunmasının istatistiksel işlemler üzerindeki

olumsuz etkilerinin araştırıldığı çalışmalar bulunmaktadır. Bu çalışmalardan bazılarında DMF'li maddelerin test eşitleme sürecine etkileri (Atalay Kabasakal, 2014; Chu, 2002; Chu ve Kamata, 2005; Turhan, 2006) ve bilgisayara adapte edilmiş testler üzerindeki etkileri (Miller,1992; Zwick, Thayer ve Wingersky, 1995; Zwick, 2000) örnek olarak verilebilir. Çalışmaların sonuçlarına göre bir testte DMF'li maddelerin yer alması doğrudan madde ve yetenek parametrelerini, dolaylı yoldan da bu parametreler ile gerçekleştirilen istatistiksel çalışmaları olumsuz etkileyebilmektedir. Yapılan bu çalışma ile de bir testin DMF'li madde içeren ve DMF'li maddeden arındırılmış formlarından kestirilen yetenek kestirimleri arasında ilişki bulunmadığı, başka bir ifadeyle DMF'li maddelerin testten çıkarılmasıyla bireylerin başarı sıralamalarının değiştiği sonucuna ulaşılmıştır. Bu sonuç ile DMF'li madde içeren bir testin DMF'li maddelerden arındırılmasının önemi ortaya konmuştur.

5.2. Öneriler

5.2.1. Araştırmacılara Yönelik Öneriler

Bu araştırma bir simülasyon çalışmasıdır. DMF üzerine gerçekleştirilen simülasyon çalışmalarının gerçek verilerle birlikte kullanılmasının daha güvenilir sonuçlar elde edilebileceği düşünülerek benzer bir çalışma gerçek verilerle desteklenen bir simülasyon çalışması ile gerçekleştirilebilir.

Bu çalışmada manipüle edilen değişkenler olarak madde sayısı, örneklem büyüklüğü ve DMF oranı kullanılmıştır. Benzer bir çalışma farklı değişkenler (Referans odak grup oranı gibi) manipüle edilerek gerçekleştirilebilir.

Bu çalışmada manipüle edilen değişkenler ikişer düzeyde ele alınmıştır. Bu değişkenlerin düzey sayıları farklılaştırılarak benzer çalışma yürütülebilir.

Bu çalışmada DMF'li maddelerin testten arındırılmasının yetenek parametrelerinin kestirimi üzerindeki etkisi araştırılmıştır. Benzer bir çalışma madde parametreleri kestirimi ile yapılabilir.

5.2.2. Uygulayıcılara Yönelik Öneriler

Araştırma sonucunda testin DMF'li maddelerden arındırılmasının bireylerin sıralamalarını deęiřtirdięi görölmüřtür. Bu sonuca göre uygulayıcıların öncelikle bir testte DMF'li madde varsa tespit edilip o maddenin uygun kořullara göre arındırılmasıyla test sonuçlarını yeniden elde etmeleri önerilebilir.



KAYNAKÇA

- Angoff, W. H. (1993). Perspectives on differential item functioning methodology. In P.W. Holland & H. Wainer (Eds.), *Differential item functioning*, 3-23. Hillside, NJ: Lawrence Erlbaum.
- Awuor, R. A. (2008). *Effect of unequal sample sizes on the power of DIF detection: An IRT based monte carlo study with sibtest and Mantel-Haenszel procedures* (Unpublished Doctorate Dissertation). Virginia Polytechnic Institute and State University, Virginia.
- Atalay, K. (2010). *PISA 2006 öğrenci anketinde yer alan tutum maddelerinin değişen madde fonksiyonu açısından incelenmesi* (Yayımlanmamış Yüksek Lisans Tezi). Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Atalay Kabasakal, K. (2014). *Değişen madde fonksiyonunun test eşitlemeye etkisi* (Yayımlanmamış Doktora Tezi). Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Atar, B. ve Kamata, A. (2011). Comparison of IRT likelihood ratio test and logistic regression DIF detection procedures. *Hacettepe University Journal of Education*, 41, 36-47.
- Atılgan, H., Kan, A. ve Doğan, N. (2011). *Eğitimde ölçme ve değerlendirme* (5. Baskı). Ankara: Anı.
- Baker, F. B., ve Kim, S. (2017). *The basics of item response theory*. Basel: Springer International Publishing.
- Başusta, N. B. (2013). *PISA 2006 fen başarı testinin madde yanlılığının kültür ve dil açısından incelenmesi* (Yayımlanmamış Yüksek Lisans Tezi). Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Bakan Kalaycıoğlu, D. ve Kelecioğlu, H. (2011). Öğrenci Seçme Sınavı'nın madde yanlılığı açısından incelenmesi. *Eğitim ve Bilim*, 36 (161), 3-12.
- Baykul, Y. (2010). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması* (2. Baskı). Ankara: Pegem Akademi.

- Camilli, G. ve Shepard, L. A. (1994). *Methods for identifying biased test items*. Hollywood: Sage.
- Chu, K. L. (2002). *Equivalent group test equating with the presence of differential item functioning* (Unpublished doctorate dissertation). The Florida State University.
- Chu, K. L., ve Kamata, A. (2005). Test equating in the presence of dif items. *Journal of Applied Measurement. Special Issue: The Multilevel Measurement Model*, 6 (3), 342-354.
- Clauser, B. ve Mazor, K. (1998). Using statistical procedures to identify differential item functioning test items. *Educational Measurement: Issue and Practice*, 17, 31-44.
- Clauser, E. B., Mazor, K., ve Hambleton, K. R. (1993). The effects of purification of the matching criterion on the identification of DIF using the Mantel-Haenszel procedure. *Applied Measurement in Education*, 6(4), 269–279.
- Corey, D. M., Dunlap P. W. ve Burke, M. J. (1998). Averaging Correlations: Expected values and bias in combined Pearson rs and Fisher's z transformations. *The Journal of General Psychology*, 125(3), 245-261, doi: 10.1080/00221309809595548
- Crocker, L. ve Algina, J. (1986). *Introduction to classical and modern test theory*. USA: Harcourt Brace Jovanovich College.
- Çepni, Z. (2011). *Değişen madde fonksiyonlarının Sibtest, Mantel Haensel, Lojistik Regresyon ve madde tepki kuramı yöntemleriyle incelenmesi* (Yayımlanmamış Doktora Tezi). Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Demir, S. (2013). *PISA 2009 matematik okuryazarlığı alt testinde bulunan maddelerin Mantel-Haenszel, Sibtest ve lojistik regresyon yöntemleri ile değişen madde fonksiyonunun incelenmesi* (Yayımlanmamış Yüksek Lisans Tezi). Bolu Abant İzzet Baysal Üniversitesi, Bolu.
- Deveci, N. A. (2008). *Üniversitelerarası kurul yabancı dil sınavının madde yanlılığı bakımından incelenmesi* (Yayımlanmamış Doktora Tezi). Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

- Doğan, N. ve Öğretmen, T. (2008). Değişen madde fonksiyonunu belirlemede Mantel Haenszel, Ki-kare ve Lojistik Regresyon tekniklerinin karşılaştırılması. *Eğitim ve Bilim*, 33, 100-112.
- Dorans, N. J., & Holland, P. W. (1993). DIF detection and description: Mantel Haenszel and standardization. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 35-66). Hillsdale, NJ: Erlbaum.
- Embretson, S. E. ve Reise, S. P. (2000). *Item response theory for psychologists*. New Jersey: Lawrence Erlbaum Associates, Inc.
- Erdem, B. (2015). *Ortaöğretime geçişte kullanılan ortak sınavların değişen madde fonksiyonu açısından kitapçık türlerine göre farklı yöntemlerle incelenmesi* (Yayımlanmamış Yüksek Lisans Tezi). Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Erkuş, A. (2011). *Davranış bilimleri için bilimsel araştırma süreci* (3. Baskı). Ankara: Seçkin.
- Ertürk, S. (1972). *Eğitimde program geliştirme*. Ankara: Yelkenetepe.
- French, B. F. ve Maller, S. J. (2007). Iterative purification and effect size use with logistic regression for differential item functioning detection. *Educational and Psychological Measurement*, 67(3), 373-393.
- Feinberg, R. A. ve Rubright D. J. (2016). Conducting simulation studies in psychometrics. *Educational Measurement: Issue and Practice*, 35(2), 36-49, doi: <https://doi.org/10.1111/emip.12111>
- Gierl, M. H., Khaliq, S. N. ve Boughton, K. (1999). *Gender differential item functioning in mathematics and science: Prevalence and policy implications*. Paper presented at the Annual Meeting of the Canadian Society for the Study of Education, Canada. Zenisky, A.L.
- Gök, B., Kelecioğlu, H. ve Doğan N. (2010). Değişen madde fonksiyonunu belirlemede Mantel-Haenszel ve Lojistik Regresyon tekniklerinin karşılaştırılması. *Eğitim ve Bilim*, 35(156).

- Hallgren, K. A. (2013). Conducting simulation studies in the R programming environment. *Tutorials in Quantitative Methods for Psychology*, 9(2), 43-60, doi: <https://doi.org/10.20982/tqmp.09.2.p043>
- Hambleton, R. K. ve Swaminathan, H. (1985). *Item response theory. Principles and applications*. Boston College: Kluwer Academic.
- Hambleton, R. K., Swaminathan, H. ve Rogers, H. J. (1991). *Fundamentals of item response theory*. USA, California: Sage.
- Han, K. T. (2008). *Impact of item parameter drift on test equating and proficiency estimates* (Unpublished Doctorate Dissertation). University of Massachusetts, Amherst.
- Holland P. W. ve Wainer H. (1993). *Differential item functioning* (s. 35-66). Lawrence Erlbaum: Hillsdale.
- Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129-145). Hillsdale, NJ: Erlbaum.
- Jodoin, G. M., & Gierl, M. J. (2001). Evaluating type I error and power rates using an effect size measure with the logistic regression procedure for DIF detection, *Applied Measurement in Education*, 14(4), 329-349, doi: 10.1207/S15324818AME1404_2
- Kamata, A. ve Vaughn, B. (2004). An introduction to differential item functioning analysis. *Learning Disabilities: A Contemporary Journal*, 2(7), 49-69.
- Karasar, N. (2016). *Bilimsel araştırma yöntemi* (30. Baskı). Ankara: Nobel.
- Kelecioğlu, H. ve Göçer Sakin, S. (2014). Geçmişten günümüze geçerlik. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*. 5(2), 1-11.
- Kim, S. ve Cohen, A. S. (2002). A comparison of linking and concurrent calibration under the graded response model. *Applied Psychological Measurement*, 26(1), 25-41.

- Köse, İ. A. (2010). *Madde tepki kuramına dayalı tek boyutlu ve çok boyutlu modellerin test uzunluğu ve örneklem büyüklüğü açısından karşılaştırılması* (Yayımlanmamış Doktora Tezi). Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Kurnaz, F. B. (2006). *Peabody resim kelime testinin madde yanlılığı açısından incelenmesi* (Yayımlanmamış Yüksek Lisans Tezi). Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Lee, H. ve Geisinger, K. F. (2016). The matching criterion purification for differential item functioning analyses in a large-scale assessment. *Educational and Psychological Measurement*, 76(1), 141-163.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749.
- Miller, T. R. (1992). *Practical considerations for conducting studies of differential item functioning in a CAT environment*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.
- Narayanan, P., ve Swaminathan, H. (1994). Performance of the Mantel-Haenszel and simultaneous item bias procedures for detecting differential. *Applied Psychological Measurement*, 18(4), 315-328. doi: <https://doi.org/10.1177%2F014662169401800403>
- Narayanan, P., ve Swaminathan, H. (1996). Identification of items that show non uniform DIF. *Applied Psychological Measurement*, 20(3), 257-274. doi: <https://doi.org/10.1177%2F014662169602000306>
- Osterlind, S. J. (1983). *Test item bias*. London: Sage.
- Öğretmen, T. (2006). *Uluslararası okuma becerilerinde gelişim projesi (PIRLS) 2001 testinin psikometrik özelliklerinin belirlenmesi: Türkiye-Amerika Birleşik Devletleri örneği* (Yayımlanmamış Doktora Tezi). Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.

- Roussos, L. A. ve Stout, W. F. (1996a). Simulation studies of the effects of sample size and studied item parameters on SIBTEST and Mantel-Haenszel Type I error performance. *Journal of Educational Measurement*, 33(2), 215-230.
- Shepard, L. A., Camilli, G. ve Williams, D. M. (1984). Validity of approximation techniques for detecting item bias. *Journal of Educational Measurement*, 22, 77-105
- Sünbül, Ö. ve Ömür Sünbül, S. (2016). Değişen madde fonksiyonunun belirlenmesinde kullanılan yöntemlerde I. tip hata ve güç çalışması. *İlköğretim Online*, 15(3), 882-897.
- Svetina, D. ve Rutkowski, L. (2014). Detecting differential item functioning using generalized logistic regression in the context of large-scale assessments. <http://www.largescaleassessmentsineducation.com/content/2/1/4>, Erişim tarihi: 17.03. 2018.
- Tekin, H. (2017). *Eğitimde ölçme ve değerlendirme* (25.Baskı). Ankara: Yargı.
- Turgut, M. F. ve Baykul, Y. (2015). *Eğitimde ölçme ve değerlendirme* (7. Baskı). Ankara: Pegem Akademi.
- Thissen, D., Steinberg, L. ve Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In P. W. Holland ve H. Wainer (Eds.), *Differential item functioning* (pp. 67-113). Hillsdale NJ: Erlbaum.
- Turhan, A. (2006). *Multilevel 2PL item response model vertical equating with the presence of differential item functioning*. Unpublished doctorate dissertation, The Florida State University
- Türk Dil Kurumu Büyük Türkçe Sözlük. http://tdk.gov.tr/index.php?option=com_bts&arama=kelime&guid=TDK.GTS.5c0d94678ec859.07771275, Erişim tarihi: 02.01.2018
- Yurdugül, H. (2003). *Ortaöğretim Kurumları Seçme ve Yerleştirme Sınavı'nın madde yanlılığı açısından incelenmesi* (Yayınlanmamış Doktora Tezi). Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.

- Yıldırım, A. (2017). *PISA 2009 okuma becerileri alanındaki maddelerin tek değişkenli ve çok değişkenli eşleştirme yöntemi ile değişen madde fonksiyonlarının incelenmesi* (Yayınlanmamış Doktora Tezi). Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistics 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.
- Zwick, R., Thayer, D. T., & Wingersky, M. (1995). Effect of Rasch calibration on ability and DIF estimation in computer-adaptive tests. *Journal of Educational Measurement*, 32, 341–363.
- Zwick R. (2000). *The assessment of differential item functioning in comput adaptive tests*. In van der Linden W. J., Glas G.A. (eds) *Computerized Adaptive Testing: Theory and Practice*. Springer, Dordrecht.