

**MADDE TAKIMLARININ
GÜVENİRLİK VE DEĞİŞEN MADDE FONKSİYONU
ÜZERİNE ETKİSİ**

**THE EFFECT OF TESTLETS
ON
RELIABILITY AND DIFFERENTIAL ITEM FUNCTIONING**

Gülşen TAŞDELEN TEKER

Hacettepe Üniversitesi

Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin

Eğitim Bilimleri Anabilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalı İçin

Öngördüğü

Doktora Tezi

olarak hazırlanmıştır.

2014

Eđitim Bilimleri Enstitüsü M¼d¼rl¼đ¼'ne,

G¼lŒen TAŒDELEN TEKER'in hazırladıđı "Madde Takımlarının G¼venirlik ve DeđiŒen Madde Fonksiyonu Üzerine Etkisi" baŒlıklı bu alıŒma j¼rimiz tarafından **Eđitim Bilimleri Anabilim Dalı, Eđitimde Ölme ve Deđerlendirme Bilim Dalı'nda Doktora Tezi** olarak kabul edilmiŒtir.


BaŒkan


Prof. Dr. Selahattin GELBAL

¼ye (DanıŒman)


Do. Dr. Nuri DOĐAN

¼ye


Prof. Dr. H¼lya KELECIOĐLU

¼ye


Do. Dr. NeŒe G¼LER

¼ye


Yrd. Do. Dr. Derya OBANOĐLU AKTAN

ONAY

Bu tez Hacettepe ¼niversitesi Lisans¼st¼ Eđitim-Öđretim ve Sınav Y¼netmeliđi'nin ilgili maddeleri uyarınca yukarıdaki j¼ri ¼yeleri tarafından ... / ... / tarihinde uygun g¼r¼lm¼Œ ve Enstit¼ Y¼netim Kurulunca ... / ... / tarihinde kabul edilmiŒtir.

Prof. Dr. Berrin AKMAN
Eđitim Bilimleri Enstit¼s¼ M¼d¼r¼

MADDE TAKIMLARININ GÜVENİRLİK VE DEĞİŞEN MADDE FONKSİYONU ÜZERİNE ETKİSİ

Gülşen TAŞDELEN TEKER

ÖZ

Bu çalışmada, yerel madde bağımlılığı gösteren madde takımları üzerinde güvenilirlik ve değişen madde fonksiyonu (DMF) analizleri yürütülmüştür. Her bir madde takımında yer alan madde sayıları birbirinden farklılık gösterdiğinden güvenilirlik analizleri dengelenmemiş desenlerde genellenebilirlik (G) kuramı yürütülerek gerçekleştirilmiştir. Değişen madde fonksiyonu analizleri ise madde takımı etkisini ele alan ve almayan yöntemler ile cinsiyet ve bölüm değişkenleri üzerinden yürütülmüştür.

Çalışmada kullanılan veri seti, bir devlet üniversitesinin Yabancı Diller Yüksek Okulu tarafından yürütülen İngilizce Yeterlik Sınavındaki altı madde takımında yer alan 20 maddeye 1500 öğrencinin verdikleri cevaplardan elde edilmiştir. Testin hem bütününe yönelik hem de cinsiyet ve bölüme göre ayrı ayrı olmak üzere betimsel istatistikler ve madde istatistikleri hesaplanmıştır.

Çalışmanın amaçlarından biri, madde takımlarından oluşan testlerin güvenilirliğe etkisini belirlemek olduğu için, madde takımı etkisinin ele alınıp alınmadığı durumlar için elde edilen güvenilirlik katsayıları karşılaştırılmıştır. Bu bağlamda yürütülen G kuramı analizleri sonucu madde takımı etkisi ele alınmadığında kestirilen G ve Phi katsayılarının madde takımı etkisi ele alındığında kestirilen G ve Phi katsayılarından daha yüksek olduğu görülmüştür. Diğer bir ifadeyle madde takımı etkisi ele alınmadığında güvenirlığın görece yüksek kestirildiği sonucuna ulaşılmıştır.

Araştırmada elde edilen sonuçlar doğrultusunda güvenirligi artırmak adına madde takımı sayısındaki artışın, madde takımlarını oluşturan madde sayısındaki artıştan daha etkili olduğu sonucuna ulaşılmıştır. Ayrıca testte yer alan madde takımlarını oluşturan madde sayıları bir madde takımından diğerine fazlasıyla değiştiğinde, bu durumun güvenirligi olumsuz etkilediği görülmüştür.

Çalışmada ayrıca, madde takımlarının deęişen madde fonksiyonuna (DMF) etkisini belirlemek adına, madde takımı etkisini ele alan ve almayan iki yöntem kullanılmış ve sonuçlar karşılaştırılmıştır. Madde takımı etkisini ele alan DMF belirleme yöntemi ile hem anlamlı düzeyde DMF gösteren madde sayısı hem de kestirilen DMF düzeyi madde takımı etkisini ele almayan yöneme oranla daha fazla çıkmıştır.

DMF gösteren maddelerin eşit yetenek düzeyindeki gruplar için farklı işlemlerinin nedenleri uzman görüşleri yardımıyla araştırılmıştır.

Anahtar sözcükler: Madde takımı, güvenilirlik, deęişen madde fonksiyonu, genellenebilirlik kuramı, çok boyutlu madde tepki kuramı, madde takımı tepki kuramı, yerel madde bağımsızlığı

Danışman: Doç. Dr. Nuri DOĞAN, Hacettepe Üniversitesi, Eğitim Bilimleri Anabilim Dalı, Eğitimde Ölçme ve Deęerlendirme Bilim Dalı

THE EFFECT OF TESTLETS ON RELIABILITY AND DIFFERENTIAL ITEM FUNCTIONING

Gülşen TAŞDELEN TEKER

ABSTRACT

In this study, reliability and differential item functioning (DIF) analysis were carried out on testlets which shows local item dependency. Since the items numbers per testlets are different from one testlet to another, the reliability analysis were carried out by using unbalanced design Generalizability (G) theory. The differential item functioning analysis were carried out by using the methods taking and not taking into account the testlet effect by means of sex and department variables.

The data set for the study has been obtained from the answers of 1500 students to six testlets composed of 20 items in English Proficiency Exam of a public university's Foreign Language School. Descriptive and item statistics were calculated separately for both the whole test and subtests which were established according to sex and department.

Since one of the aims of the study was to determine the effects of testlets on reliability, the cases in which the testlet effect was and was not taken into consideration were compared. It was seen that the estimated G and Phi coefficients were higher when the effect of testlets were not taken into account than the estimated G and Phi coefficients when the effect of testlets were taken into account as a result of the G theory analysis conducted. In other words, it was concluded that the reliability was overestimated when the testlet effect was not taken in to consideration.

According to the results of the study, it was concluded that it is more effective to increase the number of testlets rather than the items within testlets to increase the reliability. Moreover, it was seen that the differentiation of items within testlets from one to another negatively affect the reliability.

In the study also, to determine the effects of testlets on differential item functioning (DIF), two methods, one considering the testlet effect and one not, were used and the results were compared. There were more items which shows statistically significant DIF and also higher level of DIF were obtained with the DIF method

considering the testlet effect compare to the method not considering the testlet effect.

By the help of experts' opinions, the reasons of differential functioning of items for the groups of same ability level were investigated.

Keywords: Testlet, reliability, differential item functioning, generalizability theory, multidimensional item response theory, testlet response theory, local item independence

Advisor: Assoc. Prof. Dr. Nuri DOĞAN, Hacettepe University, Department of Educational Sciences, Division of Measurement and Evaluation in Education

ETİK BEYANNAMESİ

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmada,

- tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- ve bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.


Gülşen TAŞDELEN TEKER

TEŞEKKÜR

Doktora çalışmam boyunca bana rehberlik ederek değerli zamanını ve desteğini benden esirgemeyen tez danışmanım ve değerli hocam Doç. Dr. Nuri Doğan'a,

Akademik gelişimimde büyük katkıları olan değerli hocalarım Prof. Dr. Hülya KELECİOĞLU ve Prof. Dr. Selahattin GELBAL'a,

Doktora tez jürimde yer alarak değerli görüşlerini paylaşan Yrd. Doç. Dr. Derya ÇOBANOĞLU AKTAN'a,

Bu sürecin her aşamasında yanımda olan, desteğini ve en önemlisi sevgisini hissettiğim değerli hocam Doç. Dr. Neşe GÜLER'e,

Araştırmada kullandığım DMF tekniğini geliştiren ve kullanmam konusunda izin vererek analizlerle ilgili sorularımı içeren tüm maillerimi içtenlikle yanıtlayarak yol gösteren ve analizleri yapmamda büyük yardımları olan Hirotaka FUKUHARA'ya,

Yoğun çalışma temposu içinde bana destek olan sevgili iş arkadaşlarım Arş. Gör. Gülden KAYA UYANIK ve Arş. Gör. Süleyman DEMİR'e,

Doktora eğitim sürecinin her aşamasını birlikte çalışarak başardığımız sevgili arkadaşlarım Dr. Melek Gülşah EROĞLU'na ve Dr. Nagihan BOZTUNÇ ÖZTÜRK'e,

Bütün eğitim hayatım boyunca bana sonsuz güvenen, destek ve sevgilerini esirgemeyen canım annem Elmas TAŞDELEN'e, canım babam Müslüm TAŞDELEN'e, ablam Güler'e, ağabeylerim Nihat ve Cevdet'e, hayatıma neşe katan kuzucuklarım Duru ve Deniz'e,

En önemlisi canım eşim Mahmut TEKER'e,

sonsuz teşekkürler...

İÇİNDEKİLER

ÖZ	iii
ABSTRACT	v
ETİK BEYANNAMESİ.....	vii
TEŞEKKÜR.....	viii
İÇİNDEKİLER.....	ix
TABLolar DİZİNİ	xi
ŞEKİLLER DİZİNİ	xii
SİMGELER VE KISALTMALAR DİZİNİ	xii
1. GİRİŞ	1
1.1. Problem Durumu	1
1.1.1. Madde Takımı Kavramı	2
1.1.3. Genellenebilirlik Kuramı	7
1.1.4. Genellenebilirlik Kuramında Dengelenmiş ve Dengelenmemiş Desenler.....	11
1.1.5. Değişen Madde Fonksiyonu (DMF)	14
1.1.6. Madde Takımları İçin Ölçme Modelleri	18
1.1.7. Madde Takımları için Ko-değişkenli İki Faktörlü Çok Boyutlu Madde Tepki Kuramı Modeline Dayanan DMF Modeli	22
1.1.8. Ko-değişkenli İki Parametrelili Lojistik MTK Modeli	25
1.2. Araştırmanın Amacı ve Önemi.....	25
1.3. Problem Cümlesi	27
1.3.1. Alt Problemler	27
1.4. Sınırlılıklar.....	28
2. İLGİLİ ARAŞTIRMALAR	29
2.1. Madde Takımı Etkisinin Güvenirliliğe Etkisi Üzerine Yapılan Çalışmalar	29
2.2. Madde Takımı Etkisinin Değişen Madde Fonksiyonuna Etkisi Üzerine Yapılan Çalışmalar	32
2.3. Madde Takımı Etkisinin Güvenirliliğe ve Değişen Madde Fonksiyonuna Etkisi Üzerine Yapılan Çalışmalar	34
2.4. İlgili Araştırmalar Özeti.....	35
3. YÖNTEM.....	36
3.1. Araştırmanın Yöntemi	36
3.2. Çalışma Grubu	36
3.3. Verilerin Analizi.....	37
3.4. Betimsel Test İstatistikleri ve Puan Dağılımları	38
3.5. Madde İstatistikleri.....	42
4. BULGULAR VE TARTIŞMA	45
4.1. Alt Problem 1'e İlişkin Bulgular ve Yorumlar	45
4.1.1. Alt Problem 1.1'e İlişkin Bulgular ve Yorumlar	45

4.1.2. Alt Problem 1.2'ye İlişkin Bulgular ve Yorumlar	47
4.2. Alt Problem 2'ye İlişkin Bulgular ve Yorumlar	53
4.2.1. Alt Problem 2.1'e İlişkin Bulgular ve Yorumlar	53
4.2.2. Alt Problem 2.2'ye İlişkin Bulgular ve Yorumlar	56
5. SONUÇ ve ÖNERİLER	62
5.1. Sonuçlar	62
5.1.1. Alt Problem 1'e İlişkin Sonuçlar	62
5.1.2. Alt Problem 2'ye İlişkin Sonuçlar.....	63
5.2. Öneriler.....	64
5.2.1. Araştırma Sonuçlarından Çıkan Öneriler	64
5.2.2. Araştırmacılara Dönük Öneriler	64
KAYNAKÇA.....	66
EKLER DİZİNİ	70
EK-1: Çalışma Kapsamında Ele Alınan İngilizce Yeterlik Sınavı Soruları.....	71
EK-2: urGENOVA Temelli Çalışan G-String Betikleri ve Analiz Çıktıları.....	78
EK-3: WinBUGS Betikleri ve Analiz Çıktıları	81
EK-4: Örnek Uzman Kanısı Formu	87
ÖZGEÇMİŞ	96

TABLolar DİZİNİ

Tablo 1.1: İki Yüzeyli Tesadüfi Dengelenmiş $bx(m:t)$ Deseni İçin Varyans Bileşenlerinin Kestirilmesi.....	11
Tablo 1.2: İki Yüzeyli Tesadüfi Dengelenmemiş $bx(m:t)$ Deseni İçin Varyans Bileşenlerinin Kestirilmesi.....	13
Tablo 1.3: DMF Düzeyi İçin β Değerlerini Yorumlama Ölçütleri	25
Tablo 3.1: Cinsiyet ve Bölüm Puan Türüne Göre Dağılımlar	37
Tablo 3.2: Testin Betimsel İstatistikleri	38
Tablo 3.3: Testin Cinsiyet ve Bölüm Gruplarına Göre Betimsel İstatistikleri.....	40
Tablo 3.4: Madde İstatistikleri.....	42
Tablo 3.5: Cinsiyet Gruplarına Göre Madde İstatistikleri	43
Tablo 3.6: Bölümlere Göre Madde İstatistikleri.....	44
Tablo 4.1: Maddelerin Madde Takımlarına Dağılımı.....	45
Tablo 4.2: Güvenirlik Katsayıları.....	46
Tablo 4.3: Dengelenmemiş $bx(m:t)$ Desenine Ait G Çalışması Sonuçları	48
Tablo 4.4: Dengelenmemiş $bx(m:t)$ Desenine Ait K Çalışması Sonuçları	50
Tablo 4.5: Ele Alınan İki Farklı Modelde Cinsiyete Göre Kestirilen DMF Büyüklükleri ve Düzeyleri	54
Tablo 4.6: Ele Alınan İki Farklı Modelde Sayısal-Eşit Ağırlık Bölümlerine Göre Kestirilen DMF Büyüklükleri ve Düzeyleri	56
Tablo 4.7: Ele Alınan İki Farklı Modelde Sayısal-Sözel Bölümlerine Göre Kestirilen DMF Büyüklükleri ve Düzeyleri.....	59
Tablo 4.8: Ele Alınan İki Farklı Modelde Eşit Ağırlık-Sözel Bölümlerine Göre Kestirilen DMF Büyüklükleri ve Düzeyleri	60

ŞEKİLLER DİZİNİ

Şekil 1.1. Tek Biçimli DMF Örneği.....	16
Şekil 1.2. Tek Biçimli Olmayan DMF Örneği.....	17
Şekil 3.1. Test Puan Dağılım Grafiği	39
Şekil 3.2. Cinsiyete Göre Test Puan Dağılım Grafikleri	40
Şekil 3.3. Bölüme Göre Test Puan Dağılım Grafikleri.....	41

SİMGELER VE KISALTMALAR DİZİNİ

KTK: Klasik Test Kuramı

MTK: Madde Tepki Kuramı (Item Response Theory – IRT)

MTTK: Madde Takımı Tepki Kuramı (Testlet Response Theory - TRT)

ÇBMTK: Çok Boyutlu Madde Tepki Kuramı

DMF: Değişen Madde Fonksiyonu (Differential Item Functioning – DIF)

DMTF: Değişen Madde Takımı Fonksiyonu (Differential Testlet Functioning - DTF)

G-Kuramı: Genellenebilirlik Kuramı

G-Çalışması: Genellenebilirlik Çalışması

K-Çalışması: Karar Çalışması

G-Katsayısı: Genellenebilirlik Katsayısı

Phi-Katsayısı: Güvenirlik Katsayısı

1. GİRİŞ

Bu bölümde arařtırmanın problem durumu, problem cümlesi, alt problemleri ve sayıtlıları verilmiřtir.

1.1. Problem Durumu

Her yıl hem ulusal düzeyde (KPSS, YDS, ALES, YGS, LYS vb.) hem de uluslararası düzeyde (TOEFL, GRE, PISA, TIMMS, PIRLS vb.) pek çok geniş ölçekli sınav yapılmaktadır. Bu sınavlardan elde edilen sonuçlar ile bireyler ve uluslar hakkında çeřitli bilgiler edinilip önemli kararlar verilmektedir. Bu sebeple yapılan bu sınavların güvenilirlik ve geçerlik kanıtlarının sunulması vazgeçilmez bir gerekliliktir.

Testleri oluřturan birimler olarak ele alınan maddelerin yazılması, test hazırlamada en önemli adımlardan biri olarak düşünülebilir. İstenilen özellikleri ölçen ve istenilir psikometrik özelliklere sahip madde yazmak oldukça zahmetli ve özen gerektiren bir iřtir (Özçelik, 2010). Bu sebeple özellikle son 20 yıldır hem ulusal hem uluslararası çaplı olarak yürütölen pek çok standartlařtırılmıř geniş ölçekli sınavlarda madde takımı (testlet) olarak adlandırılan (Wainer ve Kiely, 1987) ortak bir uyarana baėlı olarak hazırlanan, zaman ve maliyet gibi çeřitli durumlarda saėladıkları avantajlarından veya konu alanına özėü ölçme yaklařımlarından dolayı tercih edilen madde gruplarına sıklıkla rastlanmaktadır.

Bir uyaranı okumak ve anlamak için testi alan bireyin yürüttüėü zihinsel iřlem ve harcanan zaman oldukça fazladır. Bu sebeple, böyle bir uyaranı birden fazla madde ile eřleřtirmek, birey hakkında daha fazla bilgi toplamak için gerekli zamanı düşürecektir (Bradlow, Wainer ve Wang, 1999). Bu tür madde örnekleri dilsel yeterliėin ölçüldüėü testlerde -özellikle dilbilim ve dil öğretilimi alanında okuduėunu anlama, metin çözümlemesi gibi davranıřları ölçmek amacıyla- matematik testlerinde (grafikler) ve mantıksal çözümleme sorularında sıklıkla kullanılmaktadır.

1.1.1. Madde Takımı Kavramı

Madde takımı, ilk olarak Wainer ve Kiely (1987) tarafından bilgisayar ortamında bireyselleştirilmiş test (computer adaptive testing – CAT) uygulamalarındaki kapsam etkisi ve madde sıralanışı gibi bazı problemlerin üstesinden gelmek için ortaya atılmış bir kavramdır. Bu kavram, sınavı alan bireyin daha önceden belirlenmiş bir yolu takip edeceği belirli bir sayıda ve tek bir uyaran ile ilgili bir grup ilişkili madde olarak tanımlanmıştır (Wainer ve Kiely, 1987). Wainer ve Lewis (1990) daha sonra bu tanımlı genelleştirmiş ve manipüle edilecek kadar küçük ancak kendi kapsamlarını taşıyacak/oluşturacak kadar büyük olarak tanımlamışlardır.

Lee, Brennan ve Frisbie (2000) madde takımı kavramını daha geniş bir alana uygulayabilmek için madde takımlarını test geliştirmede (Wainer ve Wang, 2000), uygulamada (Wainer ve Kiely, 1987) ve puanlamada (Wainer ve Thissen, 1996) bir ölçme birimi olarak yeniden tanımlamışlardır.

Madde takımı, tek bir ortak uyaran ile ilgili bir dizi maddeyi belirtir. Madde takımını oluşturan her bir madde ile ilgili gösterilecek performans hem genel bir yeteneğe hem de kapsam veya konu ile ilgili belirli bir yeteneğe bağlıdır (Rosenbaum, 1988; Akt. Wainer ve Lewis, 1990). Örneğin, bir madde takımında yer alan maddeler ile bir okuma parçasına, bir laboratuvar uygulamasına, bir grafik sorusuna veya karmaşık bir probleme odaklanılmış olunabilir (DeMars, 2006).

Madde takımlarının eğitim uygulamalarında kullanılmasının doğurduğu yararlı sonuçlar, bunların geniş ölçekli pek çok testte de kullanılmasına yol açmıştır. Ancak madde takımlarının belirli istatistiksel özelliklerinden dolayı, bunların tek boyutlu ölçme modellerinde uygulanması bazı olumsuz durumların ortaya çıkmasına yol açmıştır (Fukuhara ve Kamata, 2007; Yen, 1993).

Madde takımlarının kullanılmasının doğurduğu teknik kaygılardan biri yerel madde bağımlılığıdır (Local Item Dependence – LID) (Wang ve Wilson, 2005; Yen, 1993; Sireci, Thissen ve Wainer, 1991; Thissen, Steinberg ve Mooney, 1989; Wainer ve Kiely, 1987). Madde takımını oluşturan maddelerin sahip oldukları ortak uyarının,

bireylerin bu maddelere verecekleri tepkilerde de bir bağımlılık yarattığı iddia edilmektedir. Örneğin, bir okuma parçası ile ilgili daha önceden bilgi sahibi olan bir birey, hem o okuma parçası ile ilgili sorulara cevap verebilecek hem de aynı yetenek düzeyinde olan ancak o konu ile ilgili daha önce bilgi sahibi olmayan bireylere göre daha avantajlı olacaktır.

Yerel Madde Bağımsızlığı (Local Item Independence)

Herhangi bir ölçme aracında bulunan herhangi iki maddenin doğru cevaplanma olasılıklarının bireyin yetenek düzeyine bağlı olarak birbirinden bağımsız olma durumu yerel madde bağımsızlığı olarak tanımlanmaktadır (Embretson ve Reise, 2000; Hambleton ve Swaminathan, 1985). Diğer bir ifadeyle, yerel madde bağımsızlığı herhangi bir çift maddenin doğru cevaplanma olasılıklarının sadece bireyin yeteneğinden etkilenmesini belirtmektedir (Yen, 1993). Bu durum matematiksel olarak aşağıdaki gibi ifade edilebilir:

$$P(X = 1, Y = 1 | \theta) = P(X = 1 | \theta)P(Y = 1 | \theta) \quad (1)$$

Eşitlikte yer alan X ve Y aynı testte yer alan iki farklı madde olmak üzere eşitlikteki terimler aşağıdaki gibi açıklanabilir:

- $X = 1$ ve $Y = 1$: Bu iki maddenin doğru cevaplanma durumu
- $P(X = 1, Y = 1 | \theta)$: θ yetenek düzeyindeki bireyin hem X maddesini hem de Y maddesini doğru cevaplama olasılığının koşullu olasılığı
- $P(X = 1 | \theta)$: θ yetenek düzeyindeki bireyin X maddesini doğru cevaplama olasılığının koşullu olasılığı
- $P(Y = 1 | \theta)$: θ yetenek düzeyindeki bireyin Y maddesini doğru cevaplama olasılığının koşullu olasılığı

Yerel madde bağımsızlığı varsayımı bir takım ölçme teorileri için önemli bir varsayımdır ve bu sebeple yerel madde bağımlılığı (local item dependence – LID) ile sonuçlanan varsayım ihlali istenmeyen sonuçlar doğurabilir. Aşağıda varsayım ihlalinin ele alındığı bazı çalışmalar ve sonuçlarına değinilmiştir:

- Klasik test kuramı açısından bakıldığında, madde takımlarının yerel madde bağımsızlığı varsayımını ihlal etmesinden dolayı ölçmenin standart hatası

(standard error of measurement) olduğundan düşük ve bu sebeple güvenilirlik olduğundan yüksek kestirilebilir (Wainer ve Wang, 2000; Wainer ve Thissen, 1996; Wainer, 1995; Yen, 1993; Sireci, Thissen ve Wainer, 1991).

- Madde tepki kuramı (MTK) açısından incelendiğinde varsayım ihlali, testten elde edilen bilginin olduğundan yüksek ve dolayısıyla yeteneğin standart hatasının olduğundan düşük kestirilmesiyle sonuçlanmaktadır (Yen, 1993; Thissen, Steinberg ve Mooney, 1989). Bu durum ayrıca madde parametrelerinin de yanlış kestirilmesine yol açmaktadır (DeMars, 2006).
- Varsayım ihlali ayrıca madde ayıricılığının yanlış kestirilmesine yol açacaktır (Yen, 1993; Masters, 1988). Buna bağlı olarak 2- ve 3-parametrelili lojistik MTK modellerinde puanlama ağırlıkları madde ayıricılığına bağlı olduğu için yetenek kestirimleri de yanlış olacaktır (Embretson ve Reise, 2000).
- Wainer ve Wang (2000) yürüttükleri çalışmada, madde takımlarındaki bağımlılık yok sayıldığında madde güçlüğüne hala iyi bir şekilde kestirildiğini ancak şans parametresinin olduğundan düşük kestirildiğini ortaya koymuşlardır. Madde ayıricılığının ise bir testte olduğundan düşük başka bir testte ise olduğundan yüksek kestirildiğini belirtmişlerdir.
- Bradlow, Wainer ve Wang (1999) yaptıkları çalışmada, madde takımlarındaki bağımlılık modellenmediğinde, madde takımlarını oluşturan maddelerin madde ayıricılıklarının olduğundan düşük, bağımsız maddelerin madde ayıricılıklarının ise olduğundan yüksek kestirildiğini göstermişlerdir.
- Ackerman (1987) yerel madde bağımlılığı olan bir grup maddenin madde ayıricılıklarının olduğundan yüksek kestirildiğini yaptığı çalışmada ortaya koymuştur.
- Wang, Bradlow ve Wainer (2002) madde takımından kaynaklanan bağımlılık etkisinin yok sayıldığı durumlarda, yetenek ve güçlük parametrelerinin, ayıricılık ve şans parametrelerine göre daha iyi kestirildiğini göstermişlerdir.

- Yapılan bir başka arařtırmada ise, madde takımı etkisinin büyük olduđu durumlarda hem madde ayırıcılıđının hem de madde güçlüđünün ortalama mutlak hatasının, üç parametrelili lojistik modelde (3PL) madde takımı etkisinin ele alındığı durumlara kıyasla oldukça büyük çıktıđı belirlenmiştir (Glas, Wainer ve Bradlow, 2002).
- Lee, Kolen, Frisbie ve Ankenmann (2001) her bir madde takımının çoklu puanlanan bir madde gibi düşünülerek ele alınmasının, yerel madde bağımlılıđı durumunun ihmal edilmesinden ve 3PL modelinin kullanılmasından kaynaklanan yanlış parametre kestiriminden daha etkili bir durum olduđunu belirtmişlerdir. Dresher (2004) ise bir testi oluşturan bütün maddeler madde takımlarında yer alıyorsa, yetenek için elde edilen RMSE deđerinin madde bağımlılıklarının ihlal edildiđi durumlarda, madde takımlarının çoklu puanlanan madde olarak ele alındığı durumlara göre çok daha yüksek çıktıđı sonucuna varmıştır (Akt., DeMars, 2006).

1.1.2. Güvenirlik

Ölçme çalışmalarında amaç, olabildiđi kadar gerçek puanlara yakın gözlenen puanlar elde etmektir. Güvenilir, bir diđer ifadeyle gerçek puanlara yakın ölçme sonuçları ise, ancak ölçmelerdeki tesadüfi hataların azlığı ölçüsünde gerçekleşir (Baykul, 2000). Bu sebeple güvenilirlik, ölçme sonuçlarının tesadüfi hatalardan arınık olma derecesi olarak tanımlanmaktadır (Turgut ve Baykul, 2012).

Güvenirlik belirlenirken genellikle madde tabanlı güvenilirlik kestirim yöntemleri kullanılmaktadır. Bu durum eđer testler birbirinden bağımsız maddelerden oluşuyorsa bir problem teşkil etmemektedir. Ancak güvenilirliđi belirlenmek istenen test, ortak bir uyarana bađlı maddeler olarak tanımlanan madde takımlarından oluşuyorsa, geleneksel güvenilirlik belirleme yöntemlerinin yanlış kestirime yol açtığı uzun yıllardır bilinen bir gerçek olarak karşımıza çıkmaktadır (Wainer, 1995; Lee ve Frisbie, 1997; Lee ve Frisbie, 1999; Lee ve Park, 2012). Pek çok temel ölçme kitabında da madde takımlarının test güvenilirliđine etkilerine ilişkin uyarılar bulunmaktadır. Örneđin, Thorndike “Educational Measurement” kitabının 1951 basımında yer alan “Reliability” bölümünde, bir okuma parçasının maddelerinin tek

bir grup olarak ele alınmasının daha doğru güvenilirlik kestirimi sağlayacağından ve benzer şekilde Guilford (1936) basımı "Psychometric Methods" kitabında, birbirine bağımlı maddelerin (interdependent) güvenilirliği düşüreceğinden bahsetmektedir (Akt. Wainer, 1995). Bu tür maddelerin her zaman birlikte değerlendirilmesini tavsiye edip, birlikte ele alınmasının doğuracağı sonucun testteki madde sayısını azaltmakla aynı etkiyi göstereceği üzerinde durmaktadır (Akt. Wainer, 1995). Sonuç olarak alan yazın tarandığında madde takımlarından oluşan test puanlarının güvenilirliğinin, madde temelli güvenilirlik kestirim yöntemleri ile belirlendiğinde olduğundan yüksek kestirilebildiğine ilişkin pek çok çalışmaya rastlanmaktadır (Lee ve Park ,2012; Hendrickson ,2001; Lee ve Frisbie, 1997; Wainer ve Thissen, 1996; Sireci, Thissen ve Wainer, 1991).

Klasik test kuramında (KTK) ölçmeyle elde edilen puanların tutarlılığı olarak ifade edilen güvenilirlik, hatanın bağlı olduğu kaynağa göre farklılık gösterebilir. Bu sebeple ölçme sonuçlarına karışan hatalar, sadece bir değişkenlik kaynağından gelen hatalar olarak ele alınır ve bu durum KTK'nın bir sınırlılığı olarak karşımıza çıkmaktadır.

Güvenirlik için yapılabilecek yaygın tanımların birinde tekrarlı ölçümlerde benzer sonuçlar elde etmek üzerine odaklanılır. Bu tanımın temeli, eğer iç tutarlılık anlamında elde edilen güvenilirlik katsayısı uygun şekilde elde edilmişse, bu güvenilirlik eşdeğer (equivalent) formlar güvenilirliğine eşit çıkacaktır. Ancak testi oluşturan bazı maddeler bir okuma parçası veya başka bir uyarıcı materyal ile ilişkili olduğunda bu maddeler arasında bağımlılık oluşmakta ve içtutarlılık anlamındaki güvenilirlik, eşdeğer formlar arasındaki korelasyona dayalı olarak hesaplanan güvenilirliğe göre daha yüksek çıkmaktadır (Lawrence, 1995).

Eğer madde takımlarından oluşan testlerin güvenilirlikleri madde temelli olarak hesaplanan güvenilirlik belirleme yöntemleri ile hesaplandığında olduğundan yüksek kestiriliyorsa, bu durum test puanlarının yanlış yorumlanmasına yol açacaktır. Alan yazın tarandığında, bu durumun güvenilirliğin ne derecede yüksek bir kestirime yol açtığı ile ilgili ise kesin bir sonuç yer almamaktadır.

Madde tepki kuramının (MTK) en önemli varsayımlarından biri yerel madde bağımsızlığıdır. Wainer ve Lewis'e (1990) göre madde takımlarından oluşan testler genellikle yerel madde bağımsızlığı varsayımını ihlal etmektedirler. Wainer ve Lewis (1990) bu problemin üstesinden gelmede izlenebilecek üç farklı yol önermişlerdir: (1) bütün madde takımlarını sadece birer maddeden oluşacak şekilde yeniden yapılandırmak, (2) madde bağımlılıklarını görmezden gelip uygun bir model ile kestirim yapmak ve (3) madde takımlarını kendisini oluşturan maddelerle tek bir madde gibi ele almak. Üçüncü yol için, Wainer ve Thissen (1996) ile Sireci, Thissen, ve Wainer (1991) Bock'un nominal cevap modelini kullanmışlardır. Bu modele göre madde takımları çoklu puanlanan maddeler gibi ele alınmaktadır. Her bir madde takımında kaç madde varsa ele alınan madde o kadar alt kategoriye sahipmiş gibi analizler yürütülmektedir. Eğer araştırmacılar bu modelden farklı bir MTK modeli kullanırlarsa hem elde edilen sonuçlar değişecektir hem de araştırmacının MTK'nın güçlü varsayımlarını (tek boyutluluk ve yerel madde bağımsızlığı) karşılaması gerekecektir (Lee ve Frisbie, 1997).

Genellenebilirlik kuramı yukarıda belirtilen hem KTK hem de MTK tabanlı yöntemlerin kullanılması sebebiyle karşılaşılabilecek problemlerin üstesinden gelmektedir. Öncelikle pek çok hata kaynağını aynı anda ele alabilmesi ile KTK yaklaşımlarından daha doğru kestirimler yapabilmektedir. Ayrıca MTK'da olduğu gibi çok güçlü varsayımları karşılama gerekliliği olmadığından madde takımlarında güvenilirlik kestirimi amacıyla kullanılabilir. Genellenebilirlik (G) kuramı, performans ölçümlerinin güvenilirlikleri ile ilgili istatistiksel bir kuramdır. G kuramı araştırmacıya varyans analizi yoluyla farklı hata kaynaklarını bir arada kullanarak onların büyüklüklerini kestirmeyi sağlar. Bir diğer ifadeyle, klasik test kuramının (KTK) aksine G kuramı tek bir analizde, ölçmedeki hata kaynaklarının hem ayrı ayrı hem de birbirleriyle etkileşimlerinin sonucu ortaya çıkardıkları etkileri göz önünde bulundurmaktadır. Bu da G kuramının önemli bir avantajıdır (Shavelson ve Webb, 1991).

1.1.3. Genellenebilirlik Kuramı

Genellenebilirlik (G) kuramı, performans ölçümlerinin güvenilirlikleri ile ilgili istatistiksel bir kuramdır. G kuramı araştırmacıya varyans analizi yoluyla farklı hata kaynaklarını bir arada kullanarak onların büyüklüklerini kestirmeyi sağlar. Bir diğer ifadeyle, klasik test kuramının (KTK) aksine G kuramı tek bir analizde, ölçmedeki hata kaynaklarının hem ayrı ayrı hem de birbirleriyle etkileşimlerinin sonucu ortaya çıkardıkları etkileri göz önünde bulundurmaktadır. Bu da G kuramının önemli bir avantajıdır (Shavelson ve Webb, 1991).

Ölçmedeki hata kaynaklarının hem ayrı ayrı hem de birbirleriyle etkileşimlerinin sonucu ortaya çıkardıkları etkileri vardır (Thompson, 2003). KTK'nın aksine G Kuramı bu durumu da göz önünde bulundurmaktadır.

G kuramı ayrıca, KTK'dan farklı olarak uygulayıcılara, puanların güvenilirliğini farklı yorumlamalar için kestirebilmeyi sağlar. KTK sonuçlarına dayalı olarak yapılabilecek tek yorum, bireylerin birbirlerine göre durumlarını karşılaştırmada kullanılan görelî karar verebilecekleri bilgi sağlarken, G kuramı hem görelî hem de mutlak kararlar için yeterli bilgiyi aynı anda sunmaktadır (Brennan, 2001a; Shavelson ve Webb, 1991; Güler, Kaya Uyanık ve Taşdelen Teker, 2012).

G kuramında ölçme, kabul edilebilir gözlemler evreninin bir örneklemdir ve bu gözlemler verilecek kararın amacına göre değişebilir. Tek yüzeyli evrende, tek bir hata kaynağı varken sosyal bilimlerdeki ölçmelerin karmaşıklığından dolayı birden fazla yüzey içeren ölçme durumlarıyla karşılaşmak olasıdır.

Ölçme Objesi (Object of Measurement) ve Yüzeyler (Facets)

Pek çok ölçme durumunda bireyler, *ölçme objesidir*. Bir başka deyişle bireyler, istenilen kararların alınacağı ölçmenin hedefleri durumundadırlar. Bu sebeple, bireylere bağlı varyans istenilen bir durum olduğu için bireylerin doğası gereği var olan farklılıklar klasik test kuramında "hata kaynağı" olarak tanımlanmadığı gibi G kuramında da *yüzey* olarak düşünülmez (Brennan, 2001a; Shavelson ve Webb, 1991; Güler ve ark., 2012). Ölçme objesinin (bireyler) tanımladığı varyans bileşenine ilişkin değer, bireyler arasındaki farklılığı gösterir ve bu farklılık istenen bir durumdur. Ancak, unutulmamalıdır ki tüm ölçme durumlarında ölçme objesinin bireyler olması bir zorunluluk değildir, diğer değişkenler de (örneğin maddeler) ölçme objesi olabilir (Brennan, 2001a; Shavelson ve Webb, 1991). Bireyler dışında kalan ve ölçmelerde farklılıklara sebep olan diğer değişkenlik kaynakları ise *yüzey* olarak tanımlanmaktadır (Brennan, 2001a). Yüzeylere ilişkin kestirilecek varyans değerlerinin ise olabildiğince küçük çıkması istenen durumdur. Bu çalışmada ele alınan problem durumunda; bireyler ölçme objesini, maddeler ve madde takımları ise yüzeyleri oluşturmaktadır.

Çaprazlanmış ve Yuvalanmış Desenler

G kuramında sıklıkla karşılaşılan çaprazlanmış ve yuvalanmış desen kavramlarını açıklamak için “bir testte yer alan maddeleri ve bu maddeleri cevaplayan öğrencileri” örnek bir ölçme durumu olarak ele alalım. Eğer her bir öğrenci (b) testte yer alan bütün maddeleri (m) cevaplandırır ise bu desen tümüyle çaprazlanmış bir desen olacaktır ve “ bxm ” şeklinde gösterilir. Eğer farklı öğrenciler (b), farklı maddeleri (m) cevaplandırır ise bu desene de tümüyle yuvalanmış desen adı verilir. Bu desende maddeler öğrencilerde “yuvalanmıştır” ve “ $m:b$ ” şeklinde gösterilir. Daha genel bir ifadeyle, eğer bir değişkenlik kaynağının koşulları diğer bir değişkenlik kaynağının bütün koşullarıyla örtüşüyorsa çapraz (crossed) bir desen oluşuyorken, bir değişkenlik kaynağının koşulları diğer değişkenin bazı koşullarıyla örtüşüyorsa değişkenlik kaynakları yuvalanmış (nested) olarak nitelendirilmektedir (Shavelson ve Webb, 1991).

Bu çalışmada, maddelerin (m) madde takımları (t) içinde yuvalandığı ve bireylerin (b) bunlarla çaprazlandığı bir desen söz konusudur ve bu desen $bx(m:t)$ şeklinde gösterilmektedir.

Genellenebilirlik ve Karar Çalışmaları

G kuramında güvenirliliğin araştırılması iki adımda gerçekleşmektedir: genellenebilirlik (G) ve karar (K) çalışmaları. G çalışmasında araştırmacı, üzerinde ölçme yaptığı örnekleme ölçmenin evrenine genellemekle ilgilenir. *Evren*; genellikle ölçümlerin elde edildiği örneklemden çok daha kapsamlı olan ölçme durumları olarak tanımlanır. Bir testten elde edilen cevapların farklı zaman süreçleri içinde aynı kalması, bir ölçme aracının iki ya da daha fazla sayıdaki formundaki puanların tutarlılığı, alt ölçekler ya da maddeler arasındaki ilişkinin araştırılması G çalışmasına örnek durumlar olarak verilebilir. Diğer taraftan, K çalışmasında ise belirli bir amaç için karar vermek üzere G çalışmasından elde edilen bilgilerden yararlanılır. Doğru bir seçme ya da yerleştirme yapabilmek için bireyleri tanımlama, bir araştırmadaki farklı grupları karşılaştırma ya da iki veya daha fazla değişken arasındaki ilişkiyi araştırma durumları K çalışmasına örnek durumlar olarak verilebilir (Güler ve ark., 2012).

Tesadüfi ve Sabit Yüzeyler

Bir yüzeyin tesadüfi ya da sabit olarak ele alınması tamamen araştırmacının kararına bağlıdır. Örneklem büyüklüğü evren büyüklüğünden çok küçükse ve örneklem evrenden tesadüfi olarak seçilmiş veya aynı evrenden seçilebilecek başka herhangi bir örneklem ile değiştirilebilecek durumdaysa, bu tür yüzeyler *tesadüfi* olarak adlandırılırlar. Diğer taraftan, eğer araştırmacının amacı örneklemin ötesinde bir genelleme yapmak değilse, ele alınan yüzeyler *sabit* olarak tanımlanır.

Görelî ve Mutlak Model

G kuramında, ölçmenin ne derece genellenebilir olduğu eldeki verilerin karar çalışmasında nasıl kullanılacağına bağlıdır. Sosyal bilimlerdeki ölçmeler genel olarak iki şekilde kullanılır: (1) bireyleri sıralamak ve (2) bir bireyin bilgi, beceri veya tutumunu mutlak bir düzeyde belirlemek. Eğer bir öğrencinin bir sınavdan geçmesi veya kalması sınavı alan diğer öğrencilerin o sınavdaki performanslarına bağlı ise araştırmacı görelî model kullanmalı; eğer öğrencinin sınavdan geçmesi veya kalması diğer öğrencilerin o sınavdaki performanslarından bağımsız olarak sadece kendi performansına bağlı ise araştırmacı mutlak model kullanmalıdır.

Genellenebilirlik ve Phi Katsayıları

G kuramı sosyal bilimlerdeki ölçmelerde hataya sebep olan varyans kaynaklarına odaklanmasına rağmen, bir güvenilirlik katsayısı da sağlamaktadır.

G kuramında ölçme hataları, görelî kararlar ve mutlak kararlar için ayrı ayrı değerlendirilir ve bu hatalardan yola çıkılarak hesaplanan güvenilirlik katsayısı görelî ve mutlak değerlendirmeler için ayrı ayrı hesaplanabilir (Shavelson ve Webb, 1991; Brennan, 2001a). G Kuramında görelî modeller için G katsayısı elde edilirken mutlak modeller için ise Phi katsayısı elde edilir. G katsayısı, bireyin gözlenen puanından o bireyin evren puanının ne derece doğru genellendiğinin bir göstergesidir. Klasik test kuramındaki güvenilirlik katsayısı gibi, G katsayısı bireylerin puanlarındaki çeşitliliğin oranını yansıtır (Shavelson ve Webb, 1991; Brennan, 2001a; Güler, 2008). Bu katsayıların nasıl elde edildiğine

“Genellenebilirlik Kuramında Dengelenmiş ve Dengelenmemiş Desenler” bölümünde detaylı olarak değinilmiştir.

1.1.4. Genellenebilirlik Kuramında Dengelenmiş ve Dengelenmemiş Desenler

G kuramında karşılaşılan önemli bir durum ise desenlerin dengelenmiş veya dengelenmemiş olma durumudur. Dengelenmiş desenlerde gözlemlerin sayısı değişkenlik kaynağının her düzeyinde eşittir (Brennan, 2001a). Örneğin 10 birey (b) 20 madde (m) ve beş madde takımından (t) oluşan bir testi ele alalım. Bu örnekteki ölçme objesi testi alan bireyler ve yüzeyler ise maddeler ve madde takımları olmak üzere bu durumdaki desen, maddelerin madde takımlarının içinde yuvalandığı ve bireylerin bunlarla çaprazlandığı $bx(m:t)$ deseni şeklinde gösterilebilir. Bu örnekte her bir madde takımında eşit sayıda madde yer almaktadır. Diğer bir ifadeyle beş madde takımının her birinde dörder madde bulunmak suretiyle toplam 20 maddelik bir testi bütün bireylerin cevapladığı dengelenmiş bir veri yapısı bulunmaktadır. Her bir madde takımında eşit sayıda maddenin yer almasıyla oluşturulmuş dengelenmiş $bx(m:t)$ deseninde yer alan her bir varyans bileşeni varyans analizi ile Tablo 1.1’de yer alan eşitlikler yardımıyla kestirilebilir.

Tablo 1.1: İki Yüzeyli Tesadüfi Dengelenmiş $bx(m:t)$ Deseni İçin Varyans Bileşenlerinin Kestirilmesi

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Kestirilen Varyans Bileşeni
b	KT_b	n_b-1	KO_b	$\sigma^2_b = [KO_b - KO_{bt}] / n_m n_t$
t	KT_t	n_t-1	KO_t	$\sigma^2_t = [KO_t - KO_{m:t} - KO_{bt} + KO_{bm:t}] / n_b n_m$
$m:t$	$KT_{m:t}$	$n_t(n_m-1)$	$KO_{m:t}$	$\sigma^2_{m:t} = [KO_{m:t} - KO_{bm:t}] / n_b$
bt	KT_{bt}	$(n_b-1)(n_t-1)$	KO_{bt}	$\sigma^2_{bt} = [KO_{bt} - KO_{bm:t}] / n_m$
$bm:t$	$KT_{bm:t,e}$	$n_t(n_m-1)(n_b-1)$	$KO_{bm:t,e}$	$\sigma^2_{bm:t,e} = KO_{bm:t,e}$

Tablo 1.1’de yer alan varyans değerleri kullanılarak görel ve mutlak kararlarda kullanılabilir G ve Phi katsayıları elde edilebilir. Görel kararlarda kullanılmak üzere dengelenmiş desenler için kestirilen G katsayısı Eşitlik 2’deki gibi hesaplanmaktadır. Eşitlik 2’de yer alan görel hata varyansı $[\sigma^2(\delta)]$ ise Eşitlik 3 ile elde edilmektedir.

$$E\rho^2 = \frac{\sigma^2(\tau)}{\sigma^2(\tau) + \sigma^2(\delta)} \quad (2)$$

$$\sigma^2(\delta) = \frac{\sigma_{bt}^2}{n'_i} + \frac{\sigma_{bmt}^2}{n'_m n'_i} \quad (3)$$

Mutlak kararlarda kullanılmak üzere dengelenmiş desenler için kestirilen Phi katsayısı ise Eşitlik 4'teki gibi hesaplanmaktadır. Eşitlik 4'te yer alan mutlak hata varyansı $[\sigma^2(\Delta)]$ ise Eşitlik 5 ile elde edilmektedir.

$$\phi = \frac{\sigma^2(\tau)}{\sigma^2(\tau) + \sigma^2(\Delta)} \quad (4)$$

$$\sigma^2(\Delta) = \frac{\sigma_t^2}{n'_i} + \frac{\sigma_{bt}^2}{n'_i} + \frac{\sigma_{mt}^2}{n'_m n'_i} + \frac{\sigma_{bmt}^2}{n'_m n'_i} \quad (5)$$

Madde takımlarından oluşan testler incelendiğinde, madde takımlarında yer alan madde sayılarının farklılık göstermesiyle sıklıkla karşılaşılmakta ve bu durum ise *dengelenmemiş* veri setlerinin ortaya çıkmasına yol açmaktadır. Dengelenmemiş desenlerde değişken başına elde edilen gözlemler eşit değildir. Bunun sebebi olarak ise iki temel durumla karşılaşabiliriz: (1) kayıp verilerin varlığı ve (2) değişkenlerin düzeylerine ait gözlem sayısının gerçekten değişkenlik göstermesi (Brennan, 2001a).

Dengelenmemiş verilerin analizleri dengelenmiş veri yapılarına kıyasla daha karmaşıktır ve dolayısıyla dengelenmemiş veri setlerinin G kuramı analizleri daha geniş istatistiksel işlemler gerektirmektedir (Jarjoura ve Brennan, 1981). Bu sebeple dengelenmemiş veri setlerinden dengelenmiş veri setleri oluşturmak adına bazı verilerin silinmesi yoluna gidilebilmektedir. Ancak bu durumda veri kaybı söz konusu olacağından analizlerin dengelenmemiş veri setleriyle yürütülmesi avantaj sağlayacaktır (Bell, 1985).

Daha önce ele alınan maddelerin madde takımlarında yuvalandığı ve bireylerin bunlarla çaprazlandığı $bx(m:t)$ deseni, bu adımda farklı bir şekilde ele alınacaktır. Daha önce madde takımlarında eşit sayıda bulunan maddelerin sayıları farklılık gösterdiğinde ele alınan örnek, dengelenmemiş veri setine dönüşecektir. Dengelenmemiş $bx(m:t)$ deseni örneğine ait varyans bileşenleri Tablo 1.2'de yer alan eşitlikler yardımıyla kestirilebilmektedir.

Tablo 1.2: İki Yüzeyle Tesadüfi Dengelenmemiş $bx(m:t)$ Deseni İçin Varyans Bileşenlerinin Kestirilmesi

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Kestirilen Varyans Bileşeni
b	KT_b	n_b-1	KO_b	$\sigma^2_b = [KO_b - r_m KO_{bt}/t_m + (r_m - t_m) KO_{bm:t}/t_m]/n_{m+}$
t	KT_t	n_t-1	KO_t	$\sigma^2_t = [KO_{(t)} - KO_{(m:t)} - KO_{(bt)} + KO_{(bm:t)}]/n_b t_m$
$m:t$	$KT_{m:t}$	$n_{m+} - n_t$	$KO_{m:t}$	$\sigma^2_{m:t} = [KO_{(m:t)} - KO_{(bm:t)}]/n_b$
bt	KT_{bt}	$(n_b-1)(n_t-1)$	KO_{bt}	$\sigma^2_{bt} = [KO_{(bt)} - KO_{(bm:t)}]/t_m$
$bm:t$	$KT_{bm:t,e}$	$(n_b-1)(n_{m+} - n_t)$	$KO_{bm:t,e}$	$\sigma^2_{bm:t,e} = KO_{bm:t,e}$

Tablo 1.2’de verilen ve varyans bileşenlerini kestirmede kullanılan eşitliklerde geçen r_m ve t_m terimleri Eşitlik 6’da ve 7’de sırayla verilmiştir.

$$r_m = \sum_t \frac{n_{m:t}^2}{n_{m+}} \quad (6)$$

$$t_m = \frac{n_{m+} - r_m}{n_t - 1} \quad (7)$$

Dengelenmemiş Desenlerde G ve Phi Katsayısı Hesaplamaları

Dengelenmemiş $bx(m:t)$ deseni için kestirilen bağıl ve mutlak hata varyansları ve dolayısıyla G ve Phi katsayıları, dengelenmiş desen ile kestirilen değerlerden farklılık göstermektedir.

Her bir madde takımında farklı sayıda maddenin bulunduğu $bx(m:t)$ deseni için kestirilen G katsayısı, Eşitlik 8’deki gibidir. G katsayısının hesaplanması için gerekli bileşenler ise Eşitlik 9 ve 10’da verilmektedir. Bunların yanında Eşitlik 11’de yer alan \tilde{n}_t bileşeni, sadece dengelenmemiş desenlerde kullanılan bir bileşendir.

$$E\rho^2 = \frac{\sigma^2(\tau)}{\sigma^2(\tau) + \sigma^2(\delta)} \quad (8)$$

$$\sigma^2(\tau) = \sigma^2(b) \quad (9)$$

$$\sigma^2(\delta) = \frac{\sigma_{bt}^2}{\tilde{n}_t} + \frac{\sigma_{bm:t}^2}{n_{m+}} \quad (10)$$

$$\tilde{n}_t = \frac{n_{m+}^2}{\sum_t n_{m:t}^2} \quad (11)$$

Her bir madde takımında farklı sayıda maddenin bulunduğu $bx(m:t)$ deseni için kestirilen Phi katsayısı, Eşitlik 12'deki gibidir. Phi katsayısının hesaplanması için gerekli bileşenler ise Eşitlik 13 ve 14'te verilmektedir.

$$\phi = \frac{\sigma^2(\tau)}{\sigma^2(\tau) + \sigma^2(\Delta)} \quad (12)$$

$$\sigma^2(\tau) = \sigma^2(b) \quad (13)$$

$$\sigma^2(\Delta) = \frac{\sigma_t^2}{\tilde{n}_t} + \frac{\sigma_{m:t}^2}{n_{m+}} + \frac{\sigma_{bt}^2}{\tilde{n}_t} + \frac{\sigma_{bm:t}^2}{n_{m+}} \quad (14)$$

1.1.5. Değişen Madde Fonksiyonu (DMF)

Testi oluşturan madde türü ne olursa olsun, psikometrik bakımdan sahip olması gereken en önemli özellik geçerlik olarak karşımıza çıkmaktadır. Geçerlik ile ilgili karşımıza çıkan önemli tehdit unsurları ise madde ve test yanlılığıdır (Clauser ve Mazor, 1998). Dolayısıyla, teste giren farklı gruplara karşı yansızlık, testlerde karşılanması gereken bir özelliktir.

Eşit yetenek düzeyinde ancak farklı gruplarda yer alan bireylerin bir maddedeki performanslarının farklı olması değişen madde fonksiyonu (DMF) ile açıklanmaktadır. Bu durum, testle ölçülen özellik bakımından benzer olan ancak cinsiyet, sosyo-ekonomik düzey gibi değişkenler açısından birbirinden farklı alt gruplarda yer alan bireylerin, bir maddeyi doğru cevaplandırma olasılıklarının farklılaşması olarak da tanımlanabilir (Hambleton, Swaminathan ve Rogers, 1991).

DMF göstermeyen bir madde için, eşit yetenek seviyesindeki bireylerin maddeye doğru cevap verme olasılıkları, bireyler farklı gruplara ait olsalar da aynı olur. Ancak, eşit yetenek seviyelerindeki farklı gruplarda farklı madde güçlükleri gözleniyorsa madde DMF sergilemektedir şeklinde yorumlanır (Millsap ve Everson, 1993).

Madde Tepki Kuramı (MTK) üzerinden düşünülecek olursa, eğer farklı alt grupların madde karakteristik eğrileri (item characteristic curves) farklıysa DMF'nin olduğu söylenebilir (Narayanan ve Swaminathan, 1996). Diğer taraftan, iki grubun ortalama yeteneklerindeki farklılıktan kaynaklanan ve bir maddede bu grupların

farklı performans göstermeleri ise “etki (impact)” olarak adlandırılmaktadır (Dorans ve Holland, 1992). Bu sebeple, *etki*, bir maddede gösterilen performansın iki grubun gerçek yetenek farklılıklarından kaynaklanmakta iken *DMF* aynı yetenek düzeyinde o maddede farklı performans gösterilmesi olarak karşımıza çıkmaktadır. Standartlaştırılmış testlerin uygulandığı her durumda etki ile karşılaşmak mümkündür. Örneğin, bir İngilizce sınavındaki okuma parçası ile ilgili soruların cevaplanması sırasında, anadili İngilizce olan bireylerin İngilizce’yi yeni öğrenen bireylere göre daha iyi bir performans göstermesi beklenen bir durumdur.

Belirtmek gerekir ki; madde yanlılığı ile DMF birbirinden farklıdır. Diğer bir ifadeyle, DMF, yanlılığın olabileceğine işaret eden istatistiksel bir kanıt sağlayan boyutu oluşturmaktadır. Eğer madde DMF gösteriyorsa, o maddenin sosyal veya içerik açılarından daha detaylı incelenmesi gerekmektedir (Angoff, 1993). Bu inceleme genellikle alan uzmanları tarafından yapılmaktadır ve DMF gösteren maddenin gerçekten bir gruba yönelik yanlı olup olmadığı incelenir. Bu sebeple, DMF yanlı maddelerde olması gereken bir durumken, DMF’nin çıkması o maddenin kesinlikle yanlı olduğu anlamına gelmemektedir.

1.1.5.1. DMF Terminolojisi

Bu bölümde, DMF’de sıklıkla kullanılan bazı terimler üzerinde durulmuştur. Terminolojiye ek olarak, iki farklı tür DMF de, grafik gösterimi yardımıyla açıklanmıştır.

Çalışılan Maddeler (Studied Items)

Çalışılan madde (studied item, suspected item), DMF analizlerinin yürütüldüğü maddedir. Pratikte simülasyon çalışmaları dışında yürütülen analizlerde hangi maddenin DMF gösterdiği bilinmediği için, ilgilenilen bütün maddeler çalışılan maddeler olarak ele alınmaktadır.

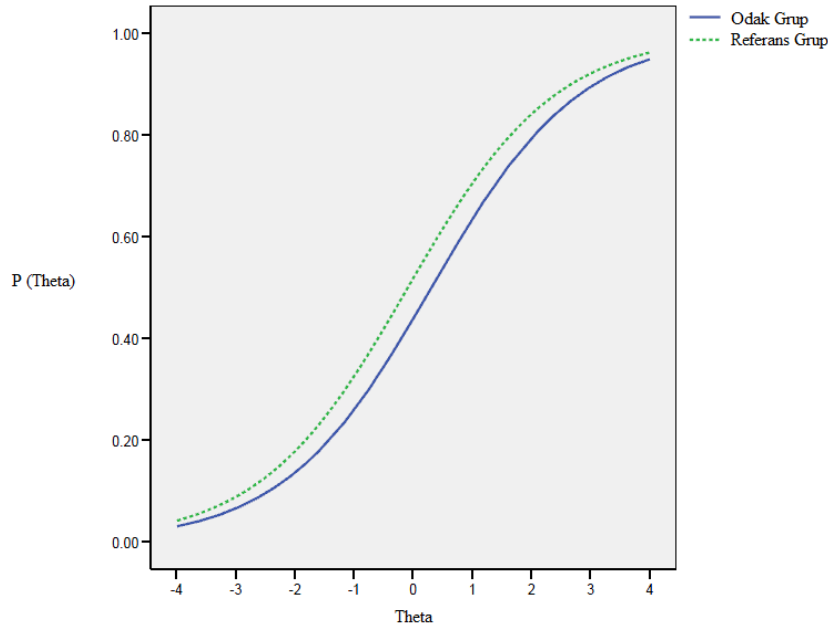
Grup Üyeliği (Group Membership)

DMF çalışmalarının odağında yer alan grup “odak grup (focal group)” olarak tanımlanmaktadır. Diğer taraftan, odak grubun karşılaştırıldığı grup ise “referans

grup (reference group)” olarak adlandırılmaktadır. Örneğin, bir İngilizce sınavının farklı gruplar için eşit olup olmadığının incelendiği bir analizde, anadili İngilizce olmayan bireylerin oluşturduğu grup odak grup, anadili İngilizce olan bireylerin oluşturduğu grup ise referans grup olarak adlandırılmaktadır.

Tek Biçimli (Uniform) ve Tek Biçimli Olmayan (Nonuniform) DMF

DMF'nin iki türü bulunmaktadır: tek biçimli ve tek biçimli olmayan DMF. Tek biçimli DMF'de, bir grup bütün yetenek düzeylerinde diğer gruptan daha yüksek veya daha düşük performans gösteriyorsa tek biçimli DMF bulunmaktadır. Tek biçimli DMF'nin bir örneği Şekil 1.1'de verilmiştir.

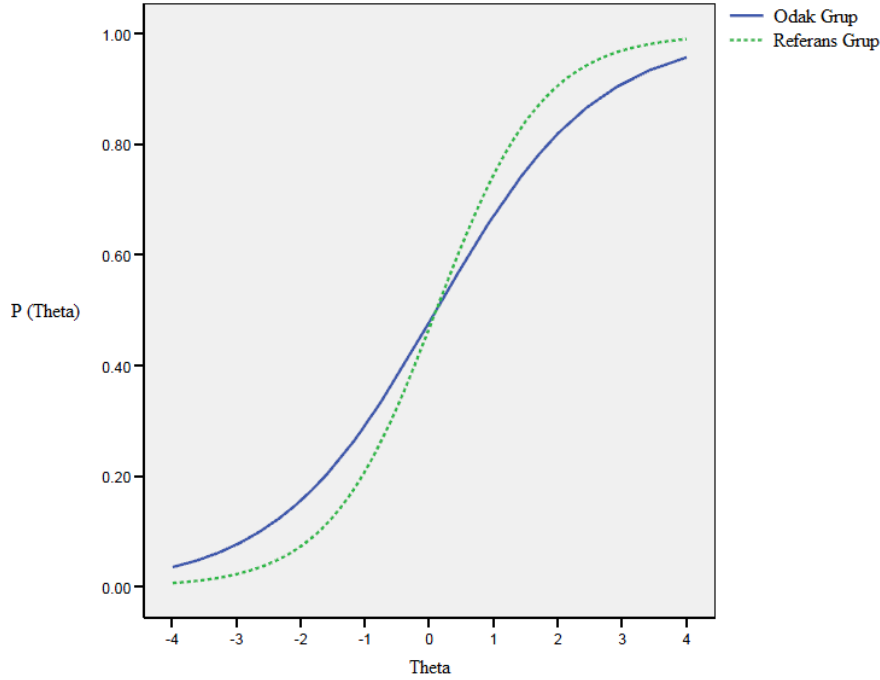


Şekil 1.1. Tek Biçimli DMF Örneği

Şekil 1.1'de hipotetik bir madde için iki madde karakteristik eğrisi verilmiştir. Bunlardan düz çizgi ile gösterilen odak grubunu, kesikli çizgiler ile gösterilen ise referans grubunu temsil etmektedir. Bu maddenin doğru cevaplandırılma olasılığı, bütün yetenek düzeyleri için referans grubunda daha yüksektir. Bu durum, tek biçimli DMF'nin kanıtıdır.

Tek biçimli olmayan DMF'de ise iki grup arasındaki farklılaşan performansın büyüklüğü ve yönü, yetenek düzeyine bağlıdır. Şekil 1.2'de tek biçimli olmayan bir

DMF örneği verilmiştir. Şekil 1.2 incelendiğinde, madde karakteristik eğrilerinin bir noktada kesiştiği görülmektedir. Maddenin doğru cevaplanma olasılığı, referans grup için bu noktanın üstündeki yetenek düzeylerinde (yaklaşık olarak 0 yetenek düzeyinin üstündeki yetenek düzeylerinde) daha yüksek iken, odak grup için bu noktanın altında (yaklaşık olarak 0 yetenek düzeyinin altındaki yetenek düzeylerinde) daha yüksektir.



Şekil 1.2. Tek Biçimli Olmayan DMF Örneği

Bazen iki madde karakteristik eğrisinin oldukça uç noktalarda (örneğin -4 veya daha düşük yetenek düzeylerinde) kesiştiği durumlara rastlanmaktadır. Bu durum, sadece bir kaç öğrencinin bu yetenek düzeyinde olmasından kaynaklanmaktadır ve bu türden tek biçimli olmayan DMF önemsenmemelidir (Fukuhara, 2009).

1.1.5.2. DMF Belirleme Yöntemleri

DMF'nin belirlenmesinde kullanılacak pek çok yöntem bulunmaktadır. Bunlar klasik test kuramına (KTK) ve madde tepki kuramına (MTK) dayalı yöntemler olarak sınıflandırılabilir. Bunlar özet şeklinde aşağıdaki başlıklarda yer alan yöntemler olarak ele alınabilir:

- **KTK Tabanlı DMF Belirleme Yöntemleri**
 - Varyans Analizi
 - Mantel-Haenszel Yöntemi
 - Lojistic Regresyon Yöntemi
 - Dönüştürülmüş Madde Güçlüğü

- **MTK Tabanlı DMF Belirleme Yöntemleri**
 - SIBTEST
 - Genelleştirilmiş Aşamalı Doğrusal Modelleme (GADM)
 - Madde Karakteristik Eğrisi
 - MIMIC
 - Lord'un Ki-kare Testi
 - Raju'nun Alan Ölçümleri
 - MTK Olabilirlik Oranı (MTK-OO)

DMF belirlemek üzere geliştirilmiş ve MTK tabanlı olup olmamalarına göre sınıflandırılan pek çok model bulunduğundan önceki bölümlerde bahsedilmişti. Ancak MTK tabanlı modeller, madde takımlarından oluşan test verilerine uygulandığında, kestirilen DMF büyüklüğü yanlı olabilmektedir (Fukuhara ve Kamata, 2007). Bu sebeple, bu tür verilerde madde takımlarından kaynaklanan yerel madde bağımlılığını göz önünde bulunduran ve bu sayede DMF büyüklüğünü daha doğru kestirebilen DMF belirleme yöntemleri tercih edilmelidir. Bu bağlamda alan yazında son yıllarda sıklıkla rastlanan madde takımları için ölçme modelleri üzerinde durulmuştur.

1.1.6. Madde Takımları İçin Ölçme Modelleri

Son yirmi yılda eğitimde kullanılan standartlaştırılmış testlerde, madde takımlarının kullanımının artması, bunların nasıl puanlanacağı ve nasıl analiz edileceğine ilişkin çalışmalarında artmasına yol açmıştır. Geleneksel olarak madde takımlarının etkisi yok sayılarak, madde takımını oluşturan her bir maddenin sanki bağımsız bir maddeymiş gibi puanlanması yoluna gidilmektedir (Lee, Kolen, Frisbie ve Ankenmann, 2001; Wainer ve Wang, 2000; Bradlow, Wainer ve Wang, 1999; Wainer, 1995; Yen, 1993; Sireci, Thissen ve Wainer, 1991). Bu çalışmalarda

yürütülen DMF analizleride, madde düzeyinde ele alınmaktadır (Wainer ve Thissen, 1996; Sireci, Thissen ve Wainer, 1991). Alan yazında madde takımlarını oluşturan maddelerin toplandığı veya ortalamalarının alınıp madde takımı düzeyinde puanların elde edildiği çalışmalara da rastlanmaktadır (Lee, Dunbar ve Frisbie, 2001; Lee ve Frisbie, 1999; Sireci, Thissen ve Wainer, 1991).

Diğer bir puanlama yöntemi olan “madde takımı puanlama”da (testlet scoring) ise madde takımları tek bir madde gibi düşünülüp çoklu puanlanmakta ve bu süreçte çoklu madde tepki modelleri -özellikle Bock’un nominal cevap modeli- kullanılmaktadır (Wainer ve Thissen, 1996; Wainer, 1995; Yen, 1993; Sireci, Wainer ve Thissen, 1991; Thissen, Steinberg ve Mooney, 1989). Ancak bu yöntemle elde edilen DMF, madde düzeyi yerine madde takımı düzeyindedir. Diğer bir ifadeyle bu yöntemle, değişen madde takımı fonksiyonu (DMTF) (differential testlet functioning - DTF) elde edilmektedir. Bu sebeple, hangi maddelerin DMF’ye yol açtığı belirlenmemektedir (Fukuhara ve Kamata, 2011). Madde takımı oluşturmak oldukça zahmetli ve zaman alıcı bir süreç olduğu için, DMTF gösteren bir madde takımının madde havuzundan çıkarılması istenen bir durum olmayacaktır. Bunun yerine madde takımında yer alan problemliler maddelerin tespit edilip düzenlenmesi, madde takımını tekrar kullanabilmek adına avantaj sağlayacaktır.

Madde takımı puanlamada, madde takımını oluşturan maddelerin birbirlerine bağlı olmalarının göz önünde bulundurulması olumlu bir durum iken bu maddeleri cevaplandıran bireyin cevap deseni ile ilgili bilgi kaybı söz konusu olmaktadır (Wainer, Bradlow ve Du, 2000). Bu olumsuzluğu ortadan kaldırmak adına, orijinal MTK modellerine kişiye özgü tesadüfi madde takımı etkisinin de eklenmesi başka bir strateji olarak ele alınmaktadır (Li, Bolt ve Fu, 2006; Wang, Bradlow ve Wainer, 2002; Wainer, Bradlow ve Du, 2000; Bradlow, Wainer ve Wang, 1999). Bu strateji ise Madde Takımı Tepki Kuramı (MTTK) (Testlet Response Theory – TRT) olarak adlandırılmaktadır (Wainer ve ark., 2000).

MTTK’da temel olarak kişiye özgü tesadüfi madde takımı etkisini ele alan pek çok madde takımı modeli bulunmaktadır (Li, Bolt ve Fu, 2006; Wang, Bradlow ve Wainer, 2002; Wainer, Bradlow ve Du, 2000; Bradlow, Wainer ve Wang, 1999).

Bütün MTTK modelleri, her bir bireydeki yerel madde bağımlılığı miktarını belirtmek için, geleneksel MTK parametrelerinin yanında bir de madde takımı parametresini önermektedir. Genel olarak, geliştirilen bütün MTTK modelleri, çok boyutlu MTK modellerinden ya da daha önce önerilen bir MTTK modelinden uyarlanmıştır.

1.1.6.1. İki- ve Üç-Parametrelili Madde Takımı Tepki Kuramı (MTTK) Modelleri

Bradlow, Wainer ve Wang (1999) iki parametrelili normal ogive modelinden uyarlanan iki parametrelili madde takımı tepki kuramı (2P-MTTK) modelini geliştirmişlerdir. Bu model iki parametrelili normal ogive modelinde var olan parametrelere ek olarak madde takımı etkisini (testlet effect) de bir parametre olarak ele almaktadır. Madde takımı etkisinden dolayı eklenen bu boyut, yerel madde bağımlılığını da göz önünde bulundurmaktadır. 2P-MTTK modeli Eşitlik 15 ile ifade edilmektedir:

$$P(y_{ij} = 1) = a_i(\theta_j - b_i + \gamma_{d(i)j}) \quad (15)$$

Eşitlik 15'te yer alan ifadeler aşağıda açıklanmıştır:

- $P(y_{ij}=1)$: j bireyinin i maddesini doğru cevaplama olasılığı
- θ_j : j bireyinin yetenek düzeyi
- a_i : i maddesinin madde ayırıcılık parametresi
- b_i : i maddesinin güçlük parametresi
- $\gamma_{d(i)j}$: j bireyi için d(i) madde takımı için tesadüfi etki

Wainer ve Wang (2000) daha önce yaptıkları çalışmalarını geliştirerek (Bradlow, Wainer ve Wang, 1999) 2P-MTTK modeline pseudo şans parametresini ekleyerek 3P-MTTK modelini önermişlerdir. Bu modelin matematiksel ifadesi ise Eşitlik 16 ile ifade edilmektedir:

$$P(y_{ij} = 1) = c_j + (1 - c_j) \logit^{-1} (a_i(\theta_j - b_i + \gamma_{d(i)j})) \quad (16)$$

Eşitlikte yer alan c_j , j maddesinin alt asimptotunu, bir diğer ifadeyle şans parametresini belirtmektedir.

1.1.6.2. Madde Takımları için İki Faktörlü Çok Boyutlu Madde Tepki Kuramı Modeli

Li, Bolt ve Fu (2006) ile DeMars (2006) çok boyutlu madde tepki kuramı perspektifinden madde takımı modelleri önermişlerdir. Madde takımları için iki faktörlü çok boyutlu madde takımı kuramı (ÇBMTK) modelinin matematiksel ifadesi Eşitlik 17'de yer almaktadır:

$$P(y_{ij} = 1) = a_{i1}\theta_j - b_i + a_{i2}\gamma_{d(i)j} \quad (17)$$

Eşitlik 17'de yer alan ifadeler aşağıda açıklanmıştır:

- θ_j ile gösterilen *yetenek* bileşeni, birinci boyutu; $\gamma_{d(i)j}$ ile gösterilen *madde takımları için tesadüfi etki* bileşeni ise ikinci boyutu temsil etmektedir. Her ikisinde standart normal dağılıma sahip olup birbirlerinden bağımsızdır.
- a_{i1} ve a_{i2} , j maddesinin ayırıcılık gücü parametreleridir. a_{i1} yeteneğin ve a_{i2} ise madde takımı etkisinin j maddesi ile arasındaki ilişkiyi belirtir. Diğer bir ifadeyle ele alınan iki boyut için ayrı ayrı ayırıcılık gücü parametreleri bulunmaktadır.
- b_i ise madde güçlüğü parametresidir.

Bu madde takımı modelinde, madde takımı etkisi 2P-MTTK ve 3P-MTTK'dan farklı olarak yorumlanmaktadır. Eğer madde takımında a_{i2} parametresi, a_{i1} parametresinden daha büyükse, madde takımlarını temsil eden ikinci boyut daha baskındır denilebilir.

Hem Li, Bolt ve Fu (2006) hem de DeMars (2006) madde takımlarının yol açtığı yerel madde bağımlılığı içeren verilerde madde takımları için iki faktörlü ÇBMTK modellerinin MTTK modellerinden daha iyi uyum sağladığı sonucuna ulaşmışlardır. Bu durumun muhtemel sebebi, madde takımları için iki faktörlü ÇBMTK modelinin birincil ve ikincil boyutlar için ayırıcılık gücünün ayrı ayrı olmasıdır. Bu sebeple, analizlerde mümkünse bu modeller tercih edilmelidir (Fukuhara ve Kamata, 2011; Li, Bolt ve Fu, 2006; DeMars, 2006).

1.1.7. Madde Takımları için Ko-değişkenli İki Faktörlü Çok Boyutlu Madde Tepki Kuramı Modeline Dayanan DMF Modeli

Daha önce değinildiği gibi araştırmacılar tarafından pek çok MTK tabanlı olan ve olmayan DMF belirleme modeli önerilmiştir. Ancak, madde takımlarından oluşan testlere MTK tabanlı DMF belirleme modelleri uygulandığında, kestirilen DMF büyüklüğü yanlış olabilmektedir (Fukuhara ve Kamata, 2007). Bu sebeple bu tür durumlarda, madde takımlarının yol açtığı yerel madde bağımlılığını da göz önünde bulunduran DMF büyüklüğünü daha doğru kestirecek modellerin kullanılması gerekmektedir.

Wang ve Wilson (2005) tarafından Rasch madde takımı modeline dayanan bir DMF belirleme modeli geliştirilmiştir. Bu model matematiksel olarak Eşitlik 18'deki gibi ifade edilebilir:

$$P(y_{ij} = 1) = \theta_j - b_i + \gamma_{d(i)j} + \beta_i G_j \quad (18)$$

Buradaki parametre yorumları Eşitlik 17'deki parametre yorumları ile benzerlik göstermektedir. Farklı olarak bu eşitlikte yer alan β_i , üzerinde çalışılan i maddesinin DMF büyüklüğünü ve G_j ise grup üyeliğini temsil etmektedir.

Eşitlik 18'deki model, madde takımı etkisini ele almakla birlikte bu model ile ilgili bir takım sınırlılıklar mevcuttur. Öncelikle model, madde ayırıcılık parametresinin testi oluşturan bütün maddeler için 1,0 olduğunu varsayan Rasch modeline dayanmaktadır. Ayrıca bu model odak ve referans grupları arasındaki ortalama yetenek düzeyi farklılıklarını ele alan bir parametre içermemektedir. Bu durum DMF büyüklüğü hesaplamada yanlışlığa yol açabilir. Bu sebeple bu çalışmada Fukuhara ve Kamata (2011) tarafından geliştirilen referans ve odak gruplar arasındaki yetenek düzeyi farklılıklarını da ele alan bir parametre içeren, madde takımları için iki faktörlü ÇBMTK modeline dayanan DMF belirleme modeli kullanılacaktır.

Fukuhara ve Kamata (2011) tarafından önerilen model temel olarak Li, Bolt ve Fu (2006) ile DeMars (2006) tarafından önerilen madde takımları için iki faktörlü çok

boyutlu madde tepki kuramı (ÇBMTK) modelinin geliştirilmiş formudur ve model matematiksel olarak Eşitlik 19'daki gibi ifade edilebilir:

$$\ln\left(\frac{P(y_{ji} = 1)}{P(y_{ji} = 0)}\right) = a_i\theta_j - \delta_i + \lambda_i\gamma_{d(i)j} - \beta'_i G_j \quad (19)$$

Eşitlik 19'daki " θ_j " birinci boyutu ya da j bireyinin yeteneğini; " $\gamma_{d(i)j}$ " ikinci boyutu ya da d madde takımının tesadüfi etkisini; " δ_i " madde güçlük parametresini; " a_i " ve " λ_i " sırasıyla yetenek ve madde takımı etkisinin ayırıcılık parametrelerini; " β'_i " DMF düzeyi belirlenmeye çalışılan i maddesinin gruplar arasındaki güçlük farkını ve " G_j " grup üyeliğini temsil etmektedir (Eğer j bireyi odak grup üyesi ise $G_j = 1$ aksi halde referans grup üyeliğinde $G_j = 0$ olacaktır). Yetenek ve madde takımlarından kaynaklanan tesadüfi etkinin madde takımları için iki faktörlü ÇBMTK modelinde olduğu gibi birbirlerinden bağımsız ve standart normal dağılıma sahip olduğu varsayılmıştır (Li, Bolt ve Fu, 2006; DeMars, 2006). Madde takımı etkisinin büyüklüğü ikinci boyutun eğiminin yeteneğin eğimine oranı olarak belirlenmiştir. i maddesinin tek biçimli (uniform) DMF düzeyinin büyüklüğü Eşitlik 19'da yer alan β'_i ile ifade edilmektedir. Sonuç olarak, i maddesi için odak grup ($G_j = 1$) için güçlük ile ilgili parametre " $\delta_i - \beta'_i$ " ile elde edilirken referans grup ($G_j = 0$) için güçlük ile ilgili parametre " δ_i " şeklindedir.

Eşitlik 19 farklı bir şekilde ifade edildiğinde Eşitlik 20 elde edilmektedir:

$$\ln\left(\frac{P(y_{ji} = 1)}{P(y_{ji} = 0)}\right) = a_i(\theta_j - b_i + C_i\gamma_{d(i)j} - \beta'_i G_j) \quad (20)$$

Eşitlik 20'deki $b_i = \frac{\delta_i}{a_i}$, $C_i = \frac{\lambda_i}{a_i}$ ve $\beta'_i = \frac{\beta'_i}{a_i}$ şeklindedir. Eşitlik 20'de yer alan $\gamma_{d(i)j}$

terimi, standart normal dağılım yerine aritmetik ortalaması sıfır ($\bar{X} = 0$) ve varyansı $\sigma_{\gamma_d}^2$ olacak şekilde yeniden ölçeklendirildiğinde Eşitlik 21'deki yeni halini almaktadır:

$$\ln\left(\frac{P(y_{ji}=1)}{P(y_{ji}=0)}\right) = a_i(\theta_j - b_i + \frac{C_i}{\sigma_{\gamma_d}}\gamma_{d(i)j} - \beta_i G_j) \quad (21)$$

Li, Bolt ve Fu (2006) eğer madde takımında yer alan bütün maddeler için $C_i = \sigma_{\gamma_d}$ olursa Eşitlik 21'in Eşitlik 22 gibi ifade edilebileceğini belirtmişlerdir:

$$\ln\left(\frac{P(y_{ji}=1)}{P(y_{ji}=0)}\right) = a_i(\theta_j - b_i + \gamma_{d(i)j} - \beta_i G_j) \quad (22)$$

Eşitlik 22'de yer alan model iki parametrelili lojistik madde takımı tepki kuramı modelinin geliştirilmiş halidir (Wang, Bradlow ve Wainer, 2002). Diğer bir ifadeyle, Eşitlik 22, yetenek ve madde takımlarından kaynaklı tesadüfi etkinin aynı madde ayırıcılık düzeyine sahip olduğunun varsayıldığı Eşitlik 19'un özel bir halidir. Bu çalışmada, Eşitlik 22'deki model ele alınmıştır. Bu sebeple bu çalışmada ele alınan ve Fukuhara ve Kamata (2011) tarafından önerilen modelin madde takımı d için $C_i = \sigma_{\gamma_d}$ olarak varsayıldığı ve madde takımları için iki faktörlü ÇBMTK modelinin geliştirilmiş hali olduğu düşünülebilir.

Ancak Eşitlik 22 odak ve referans grupları arasındaki ortalama yetenek farklılıklarını ele almamaktadır. Diğer bir ifadeyle, Eşitlik 22 odak ve referans grupları arasında yetenek düzeyi bakımından bir farklılık olmadığı varsayımına dayanmaktadır. Ancak, bu varsayımın pek çok durumda karşılanması mümkün olmamaktadır. Bu sebeple, bu tür durumlarda gruplardaki yetenek farklılıklarını ele alan bir regresyon modeli ile sorun çözülebilir (Fukuhara ve Kamata, 2011). Bu yapısal model, Eşitlik 23 ile ifade edilebilir:

$$\theta_j = \beta_\theta G_j + \zeta_j \quad (23)$$

Eşitlik 23'teki β_θ , θ_j yeteneğindeki grup etkisini (G_j) temsil etmektedir ve odak ($G_j = 1$) ve referans ($G_j = 0$) gruplar arasındaki ortalama yetenek farkını ele almaktadır. ζ_j ise, j bireyi için artık (residual) olarak ele alınmaktadır. Sonuç olarak, bu çalışmada ele alınan ve Fukuhara ve Kamata (2011) tarafından önerilen modelin matematiksel ifadesi Eşitlik 24'teki gibi ifade edilmektedir:

$$\ln\left(\frac{P(y_{ji}=1)}{P(y_{ji}=0)}\right) = a_i(\beta_\theta G_j + \zeta_j - b_i + \gamma_{d(i)j} - \beta_i G_j) \quad (24)$$

Analizler sonucu elde edilen β_i değerinin yorumlama ölçütü üç farklı düzey olarak belirlenmektedir (Vaughn, 2006). Bu ölçütler ve nasıl yorumlanacağına ilişkin bilgiler Tablo 1.3'te yer almaktadır.

Tablo 1.3: DMF Düzeyi İçin β Değerlerini Yorumlama Ölçütleri

DMF Düzeyi	Ölçüt	Yorum
A	$ \beta < 0,426$	DMF yoktur ya da ihmal edilebilir düzeydedir.
B	$0,426 \leq \beta < 0,638$	Orta düzeyde DMF vardır.
C	$ \beta \geq 0,638$	Yüksek düzeyde DMF vardır.

1.1.8. Ko-değişkenli İki Parametrelili Lojistik MTK Modeli

Bu çalışmada ele alınan madde takımları için iki faktörlü çok boyutlu madde takımı kuramı (ÇBMTK) modeline dayanan DMF belirleme modelini kontrol etmek adına, madde takımı etkisini *ele almayan* MTK tabanlı bir DMF belirleme yöntemi de kullanılmıştır. MTK tabanlı bu DMF belirleme modeli, iki parametrelili lojistik MTTK modelinin geliştirilmiş halidir. Çalışmadaki maddelerin, madde ayırıcılıkları arasında farklılıklar olabileceğinden ele alınan modelin Rasch MTTK modelinden daha iyi DMF belirleyeceği düşünülmektedir.

Bu model DMF belirlemede grup ko-değişkenlerini ele almaktadır. Modelin matematiksel ifadesi Eşitlik 25'teki gibidir:

$$\ln\left(\frac{P(y_{ji} = 1)}{P(y_{ji} = 0)}\right) = a_i(\beta_\theta G_j + \zeta_j - b_i - \beta_i G_j) \quad (25)$$

Eşitlik 25, daha önce ele alınan Eşitlik 24 ile madde takımı tesadüfi etkisi ($\gamma_{d(ij)}$) dışında birebir aynıdır. Sonuç olarak, bu çalışmanın ikinci basamağında bu iki farklı model ile madde takımlarının DMF belirlemedeki etkisi karşılaştırılmaya çalışılmıştır.

1.2. Araştırmanın Amacı ve Önemi

Bu çalışma ile yukarıda verilen bilgiler ışığında, madde takımlarından oluşan bir test kullanılarak, madde takımlarının güvenilirlik ve bir geçerlik sorunu olan değişen madde fonksiyonuna olan etkilerinin belirlenmesi amaçlanmıştır.

Son yıllarda hem ulusal hem de uluslararası düzeyde yürütülen sınavlarda madde takımlarından oluşan testlerin kullanımı artmıştır. Birbirinden bağımsız maddeler yerine madde takımlarının kullanılmasının getirdiği avantajlar da bu kullanım sıklığını olumlu bir şekilde etkilemektedir. Ancak madde takımlarının yapısından kaynaklanan yerel madde bağımlılığı durumu, bu durum göz önünde bulundurulmadığında bir takım sorunları da beraberinde getirmektedir. Bu araştırma ile ülkemizde madde takımlarından oluşan testlerde yürütülen analizlerde şimdiye kadar ele alınmayan yerel madde bağımlılığının güvenilirliğe ve değişen madde fonksiyonuna olan etkisinin ortaya konması ve söz konusu ihtiyaca cevap vermesi hedeflenmektedir.

Bu çalışmada öncelikle, madde takımından oluşan testlerin güvenilirliği kestirilken yaklaşımlar arası farklılıklar belirlenmeye çalışılmıştır. Bunun yanı sıra güvenilirliği arttırmak adına testi oluşturan madde, testte yer alan madde takımı ve madde takımlarında yer alan madde sayılarından hangisinin değiştirilmesinin güvenilirliğe en olumlu katkı sağladığı saptanmaya çalışılmıştır. Böylece madde takımlarından oluşan testlerde güvenilirliği etkileyen test yapılandırma uygulamalarının neler olacağı hakkında bilgi sağlanabilir.

Çalışmada ayrıca, madde takımı etkisini göz önünde bulunduran DMF belirleme yöntemlerinin işlevselliği ve sağladığı sonuçlar incelenmiştir. Yürütülen analizlerin yanı sıra DMF gösterdiği belirlenen maddelerin farklı işlemelerinin nedenleri uzman kanıları yoluyla irdelenmiştir.

Ülkemizde hem test güvenilirliği hem de değişen madde fonksiyonu ile ilgili yazılmış pek çok tez, yayınlanmış pek çok makale ve rapor olmasına rağmen, madde takımı etkisini ele alan bir çalışmaya rastlanmamıştır. Ulusal çapta uygulanan ALES ve YDS gibi geniş ölçekli sınavlarda sıklıkla kullanılan madde takımlarının karakteristik özelliklerinden kaynaklanan yerel madde bağımlılığı, güvenilirlik ve DMF çalışmalarını etkileyebilecek bir nitelik olduğundan, ele alınması önem taşımaktadır. Bu amaçla, bu çalışmada güvenilirlik ve DMF belirleme yöntemleri ele alınırken madde takımı etkisini göz önünde bulunduran ve bulundurmeyen yöntemler arasındaki fark ortaya konmaya çalışılmıştır.

Türkiye’de genellenabilirlik (G) kuramı ile ilgili yapılan arařtırmalar özellikle son yıllarda artış göstermekle birlikte, yürütölen arařtırmaların büyük bir bölümünde dengelenmiş desenlerin kullanıldığı görölmektedir. Nalbantođlu Yılmaz ve Uzun Bařusta (2012) ile Nalbantođlu Yılmaz (2012)’ın yürüttükleri çalıřmalar ölkemizde sınırlı sayıda yürütölmüş dengelenmemiş desenlerin kullanıldığı arařtırmalara örnek olarak gösterilebilir. Gerçek durum verilerine daha uygun olan dengelenmemiş desenlerde G kuramı analizlerinin yapılması sebebiyle bu arařtırmanın alana bir katkı sađlayacağı düşünölmektedir.

Madde takımı etkisini ele alan yöntemlerin madde parametrelerini kestirmede daha etkili olduđu yapılan pek çok çalıřma ile desteklenmiştir. Bu nedenle, madde takımlarının sıklıkla kullanıldığı hem ulusal çapta yürütölen geniş ölçekli sınavlarda hem de bu çalıřmada ele alınan türde sayıca fazla bireyin gireceđi sınavların güvenilirlik ve geçerliklerinin sorgulandıđı arařtırmalarda madde takımı etkisini göz önünde bulunduran modellerin kullanılmasına önem verilmelidir. Bu sayede oldukça kullanışlı olan madde takımlarının testlerde yer almasına rađmen elde edilen güvenilirlik kestirimi ve DMF’li maddelerin belirlenmesi daha dođru bir şekilde gerçekleştirilmiş olacaktır.

Bu dođrultuda çalıřmada ele alınan problem ařađıda bir cümle ile ifade edilmiştir.

1.3. Problem Cümlesi

Madde takımlarının güvenilirliđe ve deđişen madde fonksiyonuna etkisi var mıdır?

1.3.1. Alt Problemler

1. Madde takımlarının testin güvenilirliđi üzerindeki etkisi nasıldır?

1.1. Madde takımı etkisi dahil edilmeden madde ve madde takımı düzeylerinde kestirilen güvenilirlik katsayıları ile madde takımı etkisi dahil edilerek kestirilen güvenilirlik katsayısı arasında bir fark var mıdır?

1.2. Maddelerin (m) madde takımlarında (t) yuvalandıđı, bireylerin (b) ise bunlarla çaprazlandıđı dengelenmemiş $bx(m:t)$ deseninin genellenebilirlik kuramı sonuçları nasıldır?

1.2.1.Dengelenmemiş desende genellenebilirlik çalışması sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama oranları nasıldır?

1.2.2.Dengelenmemiş $bx(m:t)$ deseninde madde ve madde takımı sayılarının artırılıp azaltılması senaryolarına göre karar çalışmasıyla kestirilen G ve Phi katsayıları sonuçları nasıldır?

1.2.3.Testte yer alan madde takımı sayısının ve her bir madde takımında yer alan madde sayısının güvenilirlik katsayısına etkisi nasıldır?

2.Madde takımlarının değişen madde fonksiyonu üzerindeki etkisi nasıldır?

2.1.Madde takımlarının yer aldığı testlerde, madde takımı etkisinin göz önünde bulundurulup bulundurulmaması durumlarında cinsiyete göre DMF gösteren madde sayısı ve kestirilen DMF düzeyi değişmekte midir?

2.2.Madde takımlarının yer aldığı testlerde, madde takımı etkisinin göz önünde bulundurulup bulundurulmaması durumlarında bölüme göre DMF gösteren madde sayısı ve kestirilen DMF düzeyi değişmekte midir?

1.4. Sınırlılıklar

Bu araştırma, bir devlet üniversitesinin Yabancı Diller Yüksek Okulu tarafından yapılan İngilizce yeterlik sınavı uygulamasındaki “reading comprehension” testinin “text comprehension” bölümünde yer alan altı madde takımını oluşturan toplam 20 maddeyi cevaplayan 1500 bireyden elde edilen veriler ile sınırlıdır.

2. İLGİLİ ARAŞTIRMALAR

Bu bölümde madde takımı etkisinin güvenilirliğe ve değişen madde fonksiyonuna etkisini ele alan bazı araştırmalara yer verilmiştir.

2.1. Madde Takımı Etkisinin Güvenirliğe Etkisi Üzerine Yapılan Çalışmalar

Sireci, Thissen ve Wainer (1991) yaptıkları çalışmada madde takımlarından oluşan testlerde, test yapısını göz önünde bulundurmanın güvenilirlik kestiriminde yol açacağı sıkıntılar üzerinde durmuşlardır. Madde takımlarından oluşan testlerin güvenilirliklerinin belirlenmesinde MTK tabanlı bir yöntem olan Bock'un (1972) nominal cevap modelinin kullanılmasını önermektedirler. Bu amaçla çalışmada iki farklı veri seti ele alınmıştır. Her iki durumda da ele alınan verinin güvenilirliğinin kestiriminde madde ve madde takımı düzeylerinde ayrı ayrı olmak üzere hem Cronbach alfa hem de marjinal güvenilirlik katsayıları hesaplanmıştır. Maddelerin birbirlerine bağımlılıklarının göz önünde bulundurulmadığı durumda elde edilen güvenilirlik katsayılarının % 10-15 oranında daha yüksek kestirildiği tespit edilmiştir. Hatta böyle bir artışın ancak testi oluşturan madde sayısının iki katına çıkarıldığında elde edilebileceği üzerinde durulmuştur. Sonuç olarak eğer testi oluşturan maddeler yerel madde bağımsızlığına sahiplerse, madde tabanlı güvenilirlik belirleme yöntemlerinin kullanılmasının bir sakıncası bulunmamaktadır. Aksi halde yerel madde bağımlılığı gösteren madde takımlarının bir birim olarak ele alınıp güvenilirlik analizlerinin de bu şekilde yürütülmesini önermektedirler.

Wainer ve Thissen (1996) ortak bir uyarana bağlı olan maddelerde yerel madde bağımsızlığının olamayacağını ve bu bağımlılığın ise güvenilirliğin düşmesine yol açacağını belirtmişlerdir. Diğer bir ifadeyle birbirine bağımlı maddelerden oluşan bir testin güvenilirliğinin aynı sayıda bağımsız maddelerden oluşan bir teste göre daha düşük olacağını vurgulamaktadırlar. Yürütülen analizler sonucunda madde takımlarının uzunlukları arttıkça, o madde takımında yer alan maddelerin de birbirlerine bağımlılıkları artacağından güvenilirliğin düştüğü gözlenmiştir. Eğer aynı sayıda bağımsız maddeden oluşan bir testi cevaplamak için gereken süre çok daha fazlaysa, güvenilirliğin düşük çıkmasına karşın madde takımlarının kullanılması önerilmektedir.

Lee ve Frisbie (1997) çalışmalarında, daha önce yürütülen çalışmalarda madde takımlarından oluşan test puanlarının güvenilirlik kestirimlerinin, madde tabanlı güvenilirlik belirleme yöntemleri ile yapıldığında olduğundan yüksek kestirilmekten yola çıkmışlardır. Bu sebeple güvenilirlik kestiriminde madde takımı etkisini ele almayan (Cronbach alfa) ve ele alan (genellenebilirlik kuramı) yöntemleri karşılaştırmışlardır. Yürütülen analizler sonucu madde takımı etkisi ele alınmadığında kestirilen güvenirliliğin yaklaşık % 4 yüksek kestirildiğini belirlemişlerdir. Yürütülen genellenebilirlik kuramı analizleri sonucunda ise madde takımı etkisinin madde takımı – madde etkileşimi etkisinden daha etkili olduğu belirlenmiştir. Bu sebeple daha etkili ölçme araçları geliştirirken testleri oluşturan madde takımlarının sayısının değiştirilmesinin madde takımlarında yer alan madde sayılarının değiştirilmesinden daha etkili olduğu sonucuna varılmıştır.

Lee ve Frisbie (1999) diğer bir çalışmalarında, daha önce yapılan çalışmalarda madde takımlarından oluşan testlerin güvenilirlikleri madde bazında belirlendiğinde, güvenirliliğin olduğundan yüksek çıktığının belirtilmesinden dolayı, madde takımlarından oluşan testlerin güvenilirliklerinin belirlenmesinde genellenebilirlik (G) kuramının uygunluğunu araştırmışlardır. Madde düzeyinde güvenirliliğin kestirilmesinde kullanılan Cronbach alfa ile elde edilen güvenirliliğin, madde takımı etkisinin ele alınabildiği G kuramı analizlerinden daha yüksek kestirildiği sonucuna ulaşılmıştır. Çalışmada, madde takımı sayısının değiştiği ve toplam madde sayısının sabit tutulduğu durumlarda elde edilen genellenebilirlik katsayısının madde takımı sayısının sabit tutulduğu, toplam madde sayısının sabit tutulduğu ve madde takımlarındaki madde sayılarının farklılık gösterdiği duruma göre daha değişkenlik gösterdiği saptanmıştır. Bu sebeple, etkili bir ölçme yapabilmek adına madde takımı sayısının değiştirilmesinin, her bir madde takımında yer alan madde sayısının değiştirilmesinden daha etkili bir yol olduğu belirtilmiştir.

Hendrickson (2001), daha önce Lee ve Frisbie (1997) tarafından yapılan çalışmayı hem başka veriler ile tekrarlamak hem de çok boyutlu genellenebilirlik analizlerini dahil etmek adına çalışmasını yürütmüştür. Çalışmasında öncelikle, madde takımlarından oluşan testlerin güvenilirliklerini, farklı yöntemler kullanarak karşılaştırmıştır. Bu amaçla, madde ve madde takımı düzeyinde güvenilirlik belirleme yöntemleri ile tek boyutlu ve çok boyutlu genellenebilirlik kuramı

analizlerini yürüterek güvenilirlik kestirilmiştir. Alan yazında yürütülen benzer çalışmalarda olduğu gibi en yüksek güvenilirlik kestirimi madde düzeyinde ele alınan Cronbach alfa katsayısı ile elde edilmiştir. Madde takımı düzeyinde yürütülen güvenilirlik katsayısının ise olduğundan daha düşük kestirildiği belirlenmiştir. Tek boyutlu genellenebilirlik kuramı analizi sonucu elde edilen genellenebilirlik katsayısı ise madde ile madde takımı düzeylerinden elde edilen değerlerin arasında bir değer olarak kestirilmiştir. Çok boyutlu genellenebilirlik kuramı analizi sonucu elde edilen genellenebilirlik katsayısı ise tek boyutlu genellenebilirlik katsayısı ile madde düzeyinde kestirilen güvenilirlik katsayısının arasında bir değer olarak kestirilmiştir. Sonuç olarak, eğer testin özellikleri uygunsa, test ile ilgili daha çok bilgi ele alınarak analizler yürütüldüğü için, çok boyutlu genellenebilirlik kuramı analizlerinin kullanılmasının güvenilirliği daha doğru kestirebildiği sonucuna ulaşılmıştır. Çalışmada ayrıca testte yer alan madde takımı sayısının ve her bir madde takımında yer alan madde sayısının genellenebilirlik katsayısına etkisi de araştırılmıştır. Hem testte yer alan toplam madde sayısının artırılmasının hem de madde takımlarının sayısının artırılmasının güvenilirliği artırdığı sonucuna varılmıştır.

Gessaroli ve Folske (2002) yaptıkları çalışmada madde takımlarından oluşan testlerin güvenilirliklerini, madde takımlarını sabit ve tesadüfi olarak faktör analizini kullanarak belirlemişlerdir. Sonuçlar incelendiğinde, madde takımı etkisinin sabit ya da tesadüfi olarak ele alınmasının güvenilirlik kestirimlerini önemli ölçüde etkilediğini belirtmişlerdir. Bu yöntemin sıklıkla kullanılan Cronbach alfa katsayısından daha üstün bir kestirim yolu olduğunu belirtmişlerdir.

Lee ve Park (2012), alan yazında madde takımlarından oluşan testlerin güvenilirliklerinin madde tabanlı yöntemlerle olduğundan yüksek kestirildiğine yönelik çalışmaların bulunduğunu belirtmişlerdir. Bu sebeple, hangi madde tabanlı güvenilirlik kestirim yöntemlerinin güvenilirliği olduğundan yüksek kestirdiğini belirlemek ve farklı ölçme modelleri için yöntem karşılaştırmak adına bir simülasyon çalışması yürütmüşlerdir. Çalışmada güvenilirlik belirlemek adına hem genellenebilirlik (G) kuramı hem de madde tepki kuramı (MTK) için üç farklı yaklaşım kurgulanmıştır: madde puanı yaklaşımı, madde takımı puanı yaklaşımı ve maddelerin madde takımlarında yuvalandığı (madde takımı etkisinin dahil

edildiği) yaklaşım. Madde tabanlı yöntemlerle güvenilirliğin 0.02 ile 0.06 aralığında olduğundan yüksek kestirildiği görülmüştür. Madde takımı tabanlı güvenilirlik kestirimlerinin madde takımı etkisinin göz önünde bulundurulduğu durumlara göre MTK modellerindeki bilgi kaybından dolayı daha küçük kestirildiği görülmüştür. Ancak bu durum G kuramında gözlenmemiştir. Çalışmada kullanılan güvenilirlik belirleme yaklaşımlarının MTK tabanlı yöntemlerle G kuramına göre daha yüksek kestirildiği görülmüştür. MTK ve madde puanı yaklaşımlarıyla kestirilen değerlerin göreceli olarak daha küçük hatalar içerdiği sonucuna ulaşılmıştır. Bu sebeple, maddelerin madde takımlarında yuvalandığı yaklaşımın yanında madde takımı yaklaşımının hem G kuramı için hem de MTK için kullanılması sonucuna ulaşılmıştır. Ancak madde takımlarını oluşturan maddelerin yerel bağılıkları yüksekse, madde takımı etkisinin ele alındığı MTK modelinin tercih edilmesi vurgulanmıştır.

2.2. Madde Takımı Etkisinin Değişen Madde Fonksiyonuna Etkisi Üzerine Yapılan Çalışmalar

Wainer, Sireci ve Thissen (1991)'in yürüttüğü çalışmada Amerika Birleşik Devletleri'nde üniversiteye öğrenci seçiminde kullanılmakta olan Scholastic Aptitude Test (SAT) adlı sınavın deneysel versiyonunda bulunan madde takımlarını oluşturan maddelerden elde edilen veriler üzerinde değişen madde takımı fonksiyonu (differential testlet functioning) incelenmiştir. Çalışmada hem değişen madde takımı fonksiyonu tanımlanmakta hem de bu yöntemin nasıl ele alınacağı açıklanmaktadır. Çalışmada Bock (1972) tarafından geliştirilen çoklu MTK modeli kullanılmıştır. Bu modelde DMF belirlenirken öncelikle madde takımlarının hepsinin her iki grupta da (odak ve referans grupları) aynı parametrelere sahip olduğu düşünülmüştür (Madde takımlarının hiçbiri DMF göstermemektedir). Ardından bir madde takımının diğerlerinden farklı parametrelere sahip olduğu varsayılarak analizler tekrar yürütülmüştür (Madde takımlarından biri DMF göstermektedir). Her iki durumdaki olabilirlik oranları karşılaştırılmıştır. Eğer sonradan ele alınan modelde uyum anlamında olumlu bir değişme yoksa başlangıç durumunun yeterli olduğu sonucuna varılmıştır. Diğer bir ifadeyle madde takımlarının DMF göstermediği sonucuna ulaşılmıştır. Yürütülen

analizler sonucu madde takımlarından oluşan testlerde DMF belirlenirken bu yöntemin oldukça etkili olduğu vurgulanmıştır.

Wang ve Wilson (2005) madde takımı tabanlı testlerde yer alan ikili ve çoklu puanlanabilen maddelerde değişen madde fonksiyonunu (DMF) belirlemek için Rasch madde takımı modeline DMF etkisini ekleyerek çalışmalarını yürütmüşlerdir. Simülasyonlar ile yürütülen çalışmadaki bağımsız değişkenler test türü (ikili puanlanabilen maddelerden oluşan, çoklu puanlanabilen maddelerden oluşan ve ikili-çoklu puanlanabilen maddelerden oluşan testler) ve ankor madde modelidir (bir tane DMF göstermeyen madde ve tamamı DMF göstermeyen maddeler). Çalışmadaki bağımlı değişkenler ise ortalama kestirimleri arasındaki fark ve kestirilen RMSE değerleridir. Genel olarak, ankor madde sayısının çok olduğu durumlarda, DMF belirleme gücünün de daha yüksek olduğu bilinmesine rağmen, simülasyonlar sonucu hem bütün maddelerin ankor olarak ele alındığı hem de sadece bir maddenin DMF göstermediği ankor modellerinde bütün parametrelerin doğru bir şekilde elde edilebileceği bulunmuştur.

Sedivy (2009) çalışmasında madde takımlarından oluşan verilerde değişen madde fonksiyonunu incelemiştir. Bu amaçla, 3 geleneksel DMF belirleme metodu olan Mantel-Haenszel (MH), Lojistik Regresyon (LR) ve SIBTEST'in yerel bağımsızlık varsayımının ihlal edildiği madde takımı verilerindeki etkililikleri araştırılmıştır. Bunların yanında, geliştirilmiş LR ve değiştirilmiş MH yöntemleri de madde bağımlılığı ile ilgili olarak incelenmiştir. Simülasyon ve gerçek veriler üzerinde yürütülen araştırmanın sonuçlarına göre, yerel madde bağımlılığının olduğu durumlarda geliştirilmiş LR metodu daha iyi sonuç vermiştir ve bu sebeple bu metodun kullanılması önerilmektedir. Değiştirilen MH metodunun, geleneksel metoddan çok da farklı bir sonuç üretmediği belirtilmiştir. Araştırmadan elde edilen bir diğer önemli bulgu ise veri sayısı ile ilgilidir. Veri sayısı yeterince fazla olduğunda geleneksel yöntemler ile madde takımı tepki kuramı (MTTK) modelleri arasında ciddi farklılıklar gözlenmemiştir. Ancak örneklem ve DMF küçüldükçe MTTK modelleri daha iyi bir performans sergilemektedir.

Fukuhara (2009), ikili puanlanan ve madde takımlarında yer alan maddelerdeki DMF'yi tespit edebilmek için yeni bir psikometrik modeli araştırmasında önermiş ve

incelemiştir. Araştırma hem simülasyon hem de gerçek veriler üzerinden yürütülmüştür. Araştırmada önerilen model, var olan MTK modellerine göre DMF'li maddelerdeki DMF miktarını daha iyi kestirmektedir. Ayrıca önerilen model, MTK DMF modellerine göre daha yüksek bir oranda DMF tespit etmekte ve diğer parametreleri de daha iyi kestirmektedir. Sonuç olarak, hem DMF'li maddeleri daha iyi tespit edebildiği için hem de düşük ayırt ediciliğe sahip maddelerde ve düşük madde takımı etkisinde bile daha iyi sonuç verdiği için çalışmada önerilen madde takımları için iki faktörlü ÇBMTK DMF modeli, geleneksel MTK modeline tercih edilebileceği belirtilmiştir.

Fukuhara ve Kamata (2011) çalışmalarında madde takımı tabanlı veriler için DMF belirlemede kullanılabilecek bir model önermekte ve bu modeli değerlendirmektedirler. Önerilen model madde takımları için iki faktörlü çok boyutlu madde tepki kuramı modelinin geliştirilmiş halidir. DMF belirlemede kullanılan geleneksel MTK modellerinin aksine, önerilen model madde takımı etkisini de göz önünde bulundurmaktadır. Bu sayede madde takımlarından oluşan testlerde DMF'nin daha doğru kestirileceği vurgulanmaktadır. Yürütülen simülasyon çalışmaları sonucunda madde takımlarından oluşan testler için önerilen madde takımları için iki faktörlü ÇBMTK DMF belirleme modeli ile DMF büyüklüğünün belirlenmesinde daha iyi kestirimlere ulaşıldığı ve geleneksel MTK tabanlı DMF belirleme yöntemlerine göre de daha yüksek düzeyde DMF belirleme oranına sahip olduğu sonucuna ulaşılmıştır. Simülasyon çalışmalarının yanı sıra gerçek veri ile de önerilen modelin etkililiği araştırılmış ve olumlu sonuçlara ulaşılmıştır.

2.3. Madde Takımı Etkisinin Güvenirliğe ve Değişen Madde Fonksiyonuna Etkisi Üzerine Yapılan Çalışmalar

Wainer (1995), madde takımlarından oluşan testlerin, geleneksel yöntemlerde olduğu gibi bağımsız maddelermiş gibi ele alınmasının güvenilirlik kestiriminde olduğundan yüksek sonuçlar doğuracağını belirtmektedir. Bunun yanında bir geçerlik problemi olan değişen madde fonksiyonu (DMF) çalışmalarının genellikle madde bazında yürütüldüğü ancak madde takımlarından oluşan testlerin madde takımı düzeyinde incelenmesini daha doğru kestirimler sebebiyle önermektedir.

Çalışmasında da Law School Admission sınavının madde takımlarından oluşan iki farklı bölümünün hem güvenilirlik analizlerini hem de DMF analizlerinde madde takımı etkisini ele alarak yürütmüştür. Güvenirliklerin madde takımı etkisi ele alınmadığında olduğundan daha yüksek çıktığı sonucuna varmıştır. DMF sonuçlarında ise, madde düzeyinde önemsenmeyecek düzeyde çıkan DMF'nin madde takımı etkisi dahil edildiğinde belirginleştiği sonucuna varmıştır.

2.4. İlgili Araştırmalar Özeti

Yukarıda üç başlık halinde ayrı ayrı özetlenen ilgili çalışmalar gözden geçirildiğinde, yurt içinde yapılmış hiçbir çalışmaya rastlanmaması şaşırtıcı bir durumdur. Yurt dışında madde takımları üzerine uzun yıllardır çok çeşitli konularda çalışılmasına rağmen, ülkemizde madde takımlarından oluşan testlerde, madde takımı etkisini göz önünde bulundurmadan çalışmaların yürütüldüğü sonucuna ulaşılmaktadır. Yurt dışında yürütülen çalışmalar incelendiğinde ise hem güvenilirliğin daha doğru kestirilmesi hem de değişen madde fonksiyonunun daha doğru belirlenmesi amacıyla bu etkinin göz önünde bulundurulması çalışmalardan çıkarılacak genel bir sonuç olarak sunulabilir. Ancak madde takımlarının hem güvenilirliğe hem de bir geçerlik sorunu olan değişen madde fonksiyonuna etkisini bir arada ele alan çalışmaların azlığı böyle bir çalışma yapılmasında belirleyici olmuştur.

3. YÖNTEM

3.1. Araştırmanın Yöntemi

Araştırma ile madde takımlarının, testlerin güvenilirliği ve değişen madde fonksiyonu üzerine etkisini belirlemek amaçlandığı için, nicel araştırmalar kapsamında betimsel bir çalışmadır.

3.2. Çalışma Grubu

Araştırmanın çalışma grubunu, bir devlet üniversitesinin Yabancı Diller Yüksek Okulu (YDYO) İngilizce Yeterlik Sınavı (İYS) sonbahar 2011 uygulamasına katılan ve kitapçığın “reading comprehension” bölümünün “text comprehension” bölümünde yer alan 20 maddenin tümüne cevap veren adaylar oluşturmuştur.

Öğrencilerin lisans öğrenimlerinde okuyacakları bölümler sayısal, sözel ve eşit ağırlık olmak üzere üç grupta toplanarak analizlere dâhil edilmiştir. Bölümlerin kodlanması işleminde 2011 ÖSYS Kılavuzu'nda yer alan puan türleri dikkate alınmıştır.

Çalışmada İYS'de yer alan son 20 madde, madde takımlarından oluştuğu için, bu bölümdeki maddeler üzerinde çalışılmıştır. Ancak ele alınan maddelerin sınavın en sonunda yer alması sebebiyle süre, sıkılma vb sebeplerden ötürü bu sorulardan en az bir veya daha fazlasına cevap vermeyen adaylar çalışmadan çıkarılmıştır. Bu sebeple, başlangıçta 2788 olan veri sayısı eksik veri bulunan durumların çıkarılması sonucu yaklaşık %41'lik bir azalma ile 1658'e düşmüştür.

Çalışma kapsamında kullanılacak yazılımlardan biri olan G-String, en fazla 1500 kişi bulunan veri setleri ile çalışabilmektedir. Bu nedenle, çalışmada yürütülecek analizler için, SPSS yazılımı kullanılarak 1658 kişilik veriden 1500 kişilik tesadüf bir örneklem oluşturulmuştur. Çalışma kapsamında ele alınan 1500 kişilik çalışma grubunun cinsiyet ve bölümlere göre dağılımı Tablo 3.1'de verilmektedir.

Tablo 3.1: Cinsiyet ve Bölüm Puan Türüne Göre Dağılımlar

<i>Gruplar</i>		<i>N</i>	<i>%</i>
Cinsiyet	Kadın	661	44,07
	Erkek	839	55,93
Bölüm	Sayısal	1050	70
	Eşit Ağırlık	332	22,13
	Sözel	118	7,87
Toplam		1500	100

3.3. Verilerin Analizi

Çalışmada kullanılan verilerin analizinde aşağıdaki basamaklar izlenmiştir.

1. İYS’de A ve B olmak üzere iki kitapçık bulunmaktadır. Araştırmada B kitapçığını alan öğrencilerin yanıtları A kitapçığına göre düzenlenmiştir. Veriler, doğru yanıtlara 1, yanlış yanıtlara 0 puan verilerek kodlanmıştır. Çalışma kapsamında kullanılan, İYS’de yer alan “text comprehension” bölümünde yer alan 20 madde EK-1’de verilmiştir.

2. Testlerin puan dağılımları çıkarılıp, test ve madde istatistikleri ITEMAN programında hesaplanmıştır.

3. Madde takımlarından oluşan testin güvenilirliğinin belirlenmesinde, hem madde takımı etkisi dahil edilmeyerek (madde ve madde takımı düzeylerinde) hem de madde takımı etkisi dahil edilerek yürütülen genelenabilirlik kuramı analizlerinde urGENOVA (Brennan, 2001b) programı temelli çalışan G-String (Bloch ve Norman, 2011) yazılımı kullanılmıştır. Program çıktıları EK-2’de verilmiştir.

4. Madde takımlarından oluşan testlerde DMF gösteren maddelerin belirlenmesi amacıyla madde takımı etkisini ele alan “Ko-değişkenli İki Faktörlü Çok Boyutlu Madde Takımı Tepki Kuramı Modeli”ne dayanan ve madde takımı etkisini almayan “Ko-değişkenli İki Parametrelili Lojistik MTK Modeli”ne dayanan DMF belirleme yöntemleri kullanılmıştır. Bu analizler WinBUGS 1.4 (Spiegelhalter, Thomas ve Best, 2003) yazılımı kullanılarak gerçekleştirilmiştir. Program kapsamında kullanılan betikler (syntax) ve program çıktıları EK-3’te verilmiştir.

5. DMF gösteren maddelerde farklı işleyişin sebeplerine ilişkin 10 uzmandan görüş alınmıştır. Bu bağlamda görüşlerine başvurulmuş uzmanlardan biri ölçme ve değerlendirme alanında doktora sahibidir ve DMF ile ilgili çalışmaları

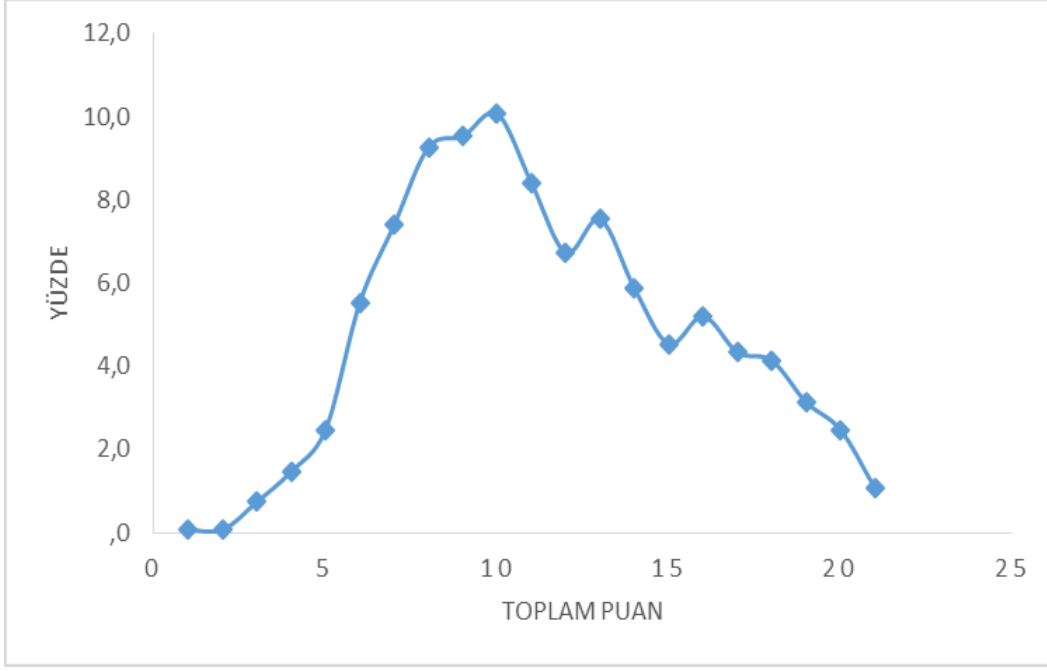
bulunmaktadır. Biri ölçme değerlendirme alanında doktora tez aşamasında öğrencidir ve DMF ile ilgili çalışmaları bulunmaktadır. Üçü ölçme değerlendirme alanında doktora tez düzeyinde öğrencidir. Diğer beş uzman ise İngilizce öğretmenliği alanından olup biri profesör, ikisi doktora düzeyinde öğrenci, ikisi ise lisans mezunu öğretmendir. Bu 10 uzmandan görüş almak adına oluşturulan örnek bir form EK-4'te verilmiştir. Özellikle İngilizce öğretmenliği alanından olan uzmanlarla yüz yüze görüşmeler yapılmış ve DMF ile ilgili gerekli açıklamalar yapıldıktan sonra görüşler toplanmıştır.

3.4. Betimsel Test İstatistikleri ve Puan Dağılımları

Çalışma kapsamında ele alınan testin betimsel istatistikleri Tablo 3.2'de, puan dağılım grafiği ise Şekil 3.1'de verilmiştir.

Tablo 3.2: Testin Betimsel İstatistikleri

<i>Değişken</i>	<i>Değer</i>
Madde sayısı	20
Öğrenci sayısı	1500
Ortalama	10,566
Standart sapma	4,089
Çarpıklık	0,242
Basıklık	-0,788
Ortalama güçlük	0,541
Ortalama ayırıcılık	0,475
Cronbach α	0,776



Şekil 3.1. Test Puan Dağılım Grafiği

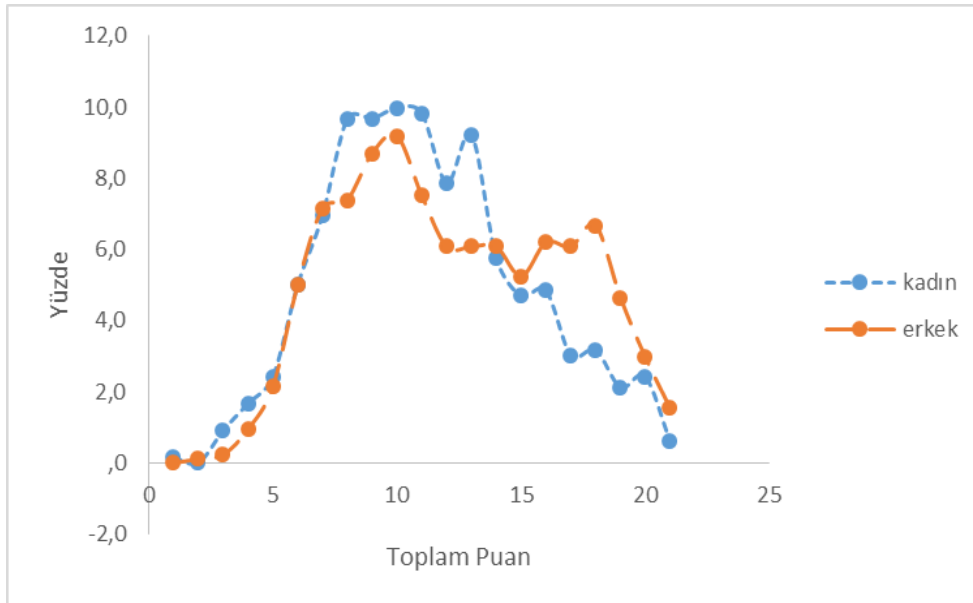
Tablo 3.2 ve Şekil 3.1 incelendiğinde testten alınan puanların dağılımının kısmen sağa çarpık ve kısmen de basık olduğu sonucuna varılabilir. Diğer yandan testin ortalama güçlükte olduğu ve maddelerin ayırıcılıklarının genelde yeterli olduğu ortalama güçlük ve ortalama ayırıcılık değerlerinden anlaşılmaktadır. Ayrıca testin güvenilirlik katsayısı da yaklaşık olarak 0,78 olarak kestirilmiştir. Bu istatistiksel özellikler testin üzerinde çalışma yapmak için bir engel bulunmadığı yönünde değerlendirilebilir.

Çalışma kapsamında ele alınan test, cinsiyet ve bölüm gruplarına göre ayrı ayrı incelendiğinde elde edilen betimsel istatistikler Tablo 3.3'te, puan dağılım grafiği ise Şekil 3.2'de verilmiştir.

Tablo 3.3: Testin Cinsiyet ve Bölüm Gruplarına Göre Betimsel İstatistikleri

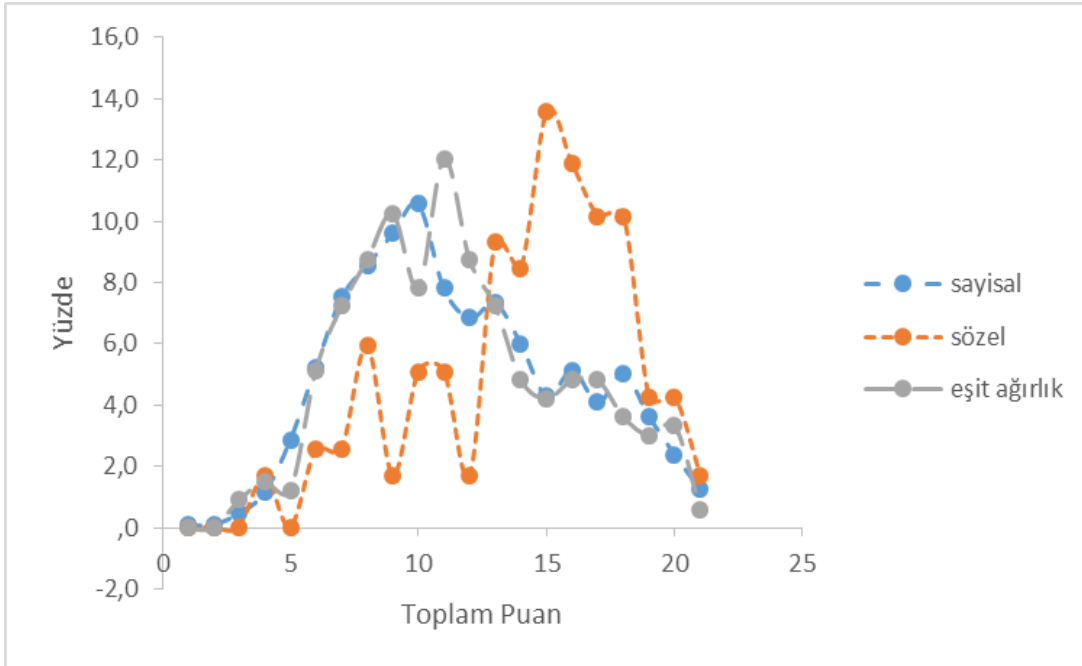
<i>Değişken</i>	<i>Cinsiyet</i>		<i>Bölüm</i>		
	<i>Kadın</i>	<i>Erkek</i>	<i>Sayısal</i>	<i>Eşit Ağırlık</i>	<i>Sözel</i>
Öğrenci sayısı	661	839	1050	332	118
Ortalama	10,230	11,242	10,264	10,151	12,636
Standart sapma	3,958	4,369	4,012	3,858	3,810
Çarpıklık	0,335	0,160	0,299	0,275	-0,513
Basıklık	-0,433	-1,027	-0,732	-0,595	-0,149
Ortalama güçlük	0,513	0,563	0,531	0,529	0,660
Ortalama ayırıcılık	0,433	0,498	0,471	0,453	0,423
Cronbach α	0,743	0,800	0,761	0,739	0,772

Tablo 3.3'te verilen betimsel istatistikler cinsiyete göre incelendiğinde erkeklerin ortalamalarının kızların ortalamalarından göreceli olarak daha yüksek olduğu görülmektedir (bir puan yüksek). Standart sapma değerleri ise birbirine yakın kestirilmiştir. Yani grupların heterojenliği birbirine yakındır. Bu durumun destekleyicisi olarak her iki grup için de negatif olarak kestirilen basıklık katsayıları da kanıt olarak gösterilebilir. Testten elde edilen ortalama güçlük ve ayırt edicilik değerleri ise birbirine çok yakın kestirilmiştir. Hem şekil 3.2'de yer alan grafik incelendiğinde hem de Tablo 3.3'te yer alan basıklık, çarpıklık, ortalama gibi değerler değerlendirildiğinde, cinsiyet gruplarına göre dağılımın birbirine yakın olduğu görülmektedir.



Şekil 3.2. Cinsiyete Göre Test Puan Dağılım Grafikleri

Tablo 3.3'te verilen betimsel istatistikler bölüme göre incelendiğinde göreceli olarak en yüksek ortalama değeri sözel bölümde okunmaktadır. Standart sapma değerleri ise her üç bölüm türünde de birbirine çok yakın kestirilmiştir. Testin ortalama güçlüğü incelendiğinde bu değer en yüksek sözel bölümde okunurken sayısal ve eşit ağırlık bölümleri için kestirilen ortalama güçlük değerleri birbirine çok yakın çıkmıştır. Testten elde edilen ayırt edicilik değeri ise üç bölüm için birbirine çok yakın kestirilmiştir. Şekil 3.3'te yer alan grafik incelendiğinde ise sayısal ve eşit ağırlık gruplarının dağılımlarının birbirine oldukça benzediği ancak sözel bölüm dağılımının farklılaştığı görülmektedir. Sayısal ve eşit ağırlık bölümleri için dağılımlar az miktarda sağa çarpık iken sözel bölüm için bu durum sola çarpık olarak görülmektedir. Bu durum Tablo 3.3'te okunan çarpıklık değerlerinden de anlaşılmaktadır. Bu durum testin ortalama değerinin sözel bölüm için daha yüksek kestirilmesi ile paralel olarak testin sözel öğrencilere daha kolay geldiği şeklinde yorumlanabilir.



Şekil 3.3. Bölüme Göre Test Puan Dağılım Grafikleri

3.5. Madde İstatistikleri

Çalışma kapsamında ele alınan testin madde istatistikleri Tablo 3.4'te verilmiştir.

Tablo 3.4: Madde İstatistikleri

<i>Madde No</i>	<i>p</i>	<i>r_{çs}</i>
1	0,376	0,604
2	0,365	0,592
3	0,644	0,494
4	0,759	0,508
5	0,575	0,556
6	0,635	0,556
7	0,617	0,422
8	0,651	0,614
9	0,389	0,484
10	0,340	0,230
11	0,350	0,424
12	0,664	0,580
13	0,513	0,648
14	0,761	0,322
15	0,717	0,478
16	0,371	0,410
17	0,487	0,396
18	0,372	0,254
19	0,555	0,506
20	0,677	0,424

Tablo 3.4'te yer alan madde güçlük indeksleri (p) (0,34 - 0,76) aralığında değerler almıştır. Madde güçlüklerine dair veriler incelendiğinde, testte zor ve kolay maddeler nispeten az olmakla birlikte orta güçlükteki maddelerin çoğunlukta olduğu gözlenmektedir. Madde ayırıcılık indeksleri ($r_{çs}$) ise (0,23 - 0,65) aralığında değerler almıştır. Madde ayırt edicilik indeksleri incelendiğinde, okunan değerler, üç madde (10, 14 ve 18. maddeler) dışında 0,39 ve üstündedir. Madde ayırıcılıkları göz önüne alındığında testte yer alan maddelerin genel olarak yüksek ayırıcılık gücüne sahip olduğu söylenebilir.

Çalışma kapsamında ele alınan sınavın cinsiyet gruplarına göre madde istatistikleri Tablo 3.5'te verilmiştir.

Tablo 3.5: Cinsiyet Gruplarına Göre Madde İstatistikleri

Madde No	Kadın		Erkek	
	<i>p</i>	<i>r_{çs}</i>	<i>p</i>	<i>r_{çs}</i>
1	0,334	0,518	0,409	0,677
2	0,303	0,459	0,414	0,670
3	0,641	0,455	0,646	0,523
4	0,731	0,491	0,782	0,491
5	0,536	0,482	0,607	0,624
6	0,648	0,536	0,625	0,581
7	0,610	0,423	0,623	0,444
8	0,623	0,614	0,672	0,599
9	0,354	0,391	0,416	0,516
10	0,333	0,273	0,346	0,176
11	0,319	0,309	0,374	0,516
12	0,641	0,523	0,682	0,599
13	0,481	0,523	0,539	0,695
14	0,717	0,309	0,795	0,315
15	0,700	0,468	0,731	0,466
16	0,322	0,327	0,410	0,452
17	0,461	0,273	0,508	0,513
18	0,349	0,264	0,390	0,219
19	0,513	0,500	0,588	0,509
20	0,638	0,532	0,708	0,384

Tablo 3.5'te verilen madde güçlükleri incelendiğinde cinsiyete göre maddeler arası çok büyük bir farklılaşma gözlenmemektedir. Madde güçlük indeksleri (*p*) kadınlar için (0,30 - 0,73) aralığında iken erkekler için (0,35 - 0,79) aralığında değerler almıştır. Güçlük düzeyleri daha detaylı incelendiğinde 6.madde dışında bütün maddeler erkekler için daha kolay gelmiştir. Benzer durum birkaç madde dışında ayırt edicilik indekslerinde de gözlenmektedir. Madde ayırtıcılık indeksleri (*r_{çs}*) kadınlar için (0,26 - 0,61) aralığında iken erkekler için (0,18 - 0,69) aralığında değerler almıştır. Madde ayırtıcılık düzeyleri daha detaylı incelendiğinde ise 8, 10, 15, 18 ve 20.maddeler dışında bütün maddelerin ayırt edicilik indeksleri erkeklerde daha yüksek kestirilmiştir.

Çalışma kapsamında ele alınan sınavın bölüm gruplarına göre madde istatistikleri ise Tablo 3.6'da verilmiştir.

Tablo 3.6: Bölümlere Göre Madde İstatistikleri

Madde No	Sayısal		Eşit Eğırlık		Sözel	
	<i>p</i>	<i>r_{çs}</i>	<i>p</i>	<i>r_{çs}</i>	<i>p</i>	<i>r_{çs}</i>
1	0,350	0,580	0,352	0,545	0,669	0,667
2	0,349	0,574	0,349	0,536	0,551	0,538
3	0,627	0,506	0,648	0,364	0,788	0,436
4	0,749	0,500	0,759	0,518	0,856	0,308
5	0,570	0,554	0,542	0,518	0,712	0,462
6	0,617	0,551	0,627	0,564	0,814	0,308
7	0,593	0,420	0,633	0,345	0,788	0,410
8	0,631	0,626	0,657	0,545	0,805	0,359
9	0,387	0,474	0,337	0,445	0,551	0,487
10	0,338	0,223	0,340	0,200	0,356	0,308
11	0,342	0,403	0,346	0,391	0,432	0,487
12	0,653	0,606	0,642	0,518	0,822	0,436
13	0,499	0,623	0,491	0,545	0,703	0,769
14	0,752	0,311	0,774	0,373	0,797	0,231
15	0,693	0,471	0,738	0,455	0,873	0,308
16	0,378	0,457	0,361	0,345	0,339	0,231
17	0,473	0,414	0,509	0,391	0,551	0,308
18	0,389	0,243	0,325	0,345	0,356	0,462
19	0,561	0,491	0,503	0,555	0,644	0,590
20	0,672	0,391	0,654	0,555	0,788	0,359

Tablo 3.6'da yer alan madde güçlüklerine dair veriler incelendiğinde sadece iki madde (16 ve 18.maddeler) dışında kestirilen değerlerin, sözel bölüm için diğer bölümlere göre daha yüksek kestirildiği gözlenmektedir. Sayısal ve eşit ağırlık bölümleri için ise madde güçlükleri incelendiğinde maddeler arası çok büyük bir farklılaşma gözlenmemektedir. Madde güçlük indeksleri (*p*) sayısal bölüm için (0,34 - 0,75) aralığında, eşit ağırlık bölümü için (0,33 - 0,77) aralığında ve sözel bölüm için (0,37 - 0,87) aralığında değerler almıştır. Madde ayırt edicilik indeksleri (*r_{çs}*) ise sayısal bölüm için (0,22 - 0,62) aralığında, eşit ağırlık bölümü için (0,20 - 0,56) aralığında ve sözel bölüm için (0,23 - 0,77) aralığında değerler almıştır. Bölümler ayırt edicilik değerleri bakımından karşılaştırıldığında sekiz madde dışında sözel bölüm için elde edilen değerlerin görece daha düşük olduğu gözlenmektedir.

4. BULGULAR VE TARTIŞMA

Bu bölümde, alt problem sırasına göre verilmiş araştırma bulguları ve bu bulgularla ilgili değerlendirmeler yer almaktadır.

4.1. Alt Problem 1'e İlişkin Bulgular ve Yorumlar

Alt problem 1. *Madde takımlarının testin güvenilirliği üzerindeki etkisi nasıldır?*

Çalışma kapsamında ele alınan ve madde takımlarından oluşan veri seti üzerinde güvenilirlik belirleme çalışmaları yürütülmüştür. Çalışma kapsamında madde takımlarından oluşan testlerin güvenilirliğe etkisi belirlenmek amaçlandığı için madde takımı etkisinin ele alındığı ve alınmadığı durumlar karşılaştırılmıştır. Güvenirlik kestirimi sırasında madde takımı etkisinin de dahil edilebildiği genellenebilirlik (G) kuramı ile analizler gerçekleştirilmiştir.

4.1.1. Alt Problem 1.1'e İlişkin Bulgular ve Yorumlar

Alt problem 1.1. *Madde takımı etkisi dâhil edilmeden madde ve madde takımı düzeylerinde kestirilen güvenilirlik katsayıları ile madde takımı etkisi dâhil edilerek kestirilen güvenilirlik katsayıları arasında bir fark var mıdır?*

Çalışma kapsamında, İngilizce yeterlik sınavına giren 1500 öğrencinin altı madde takımında yuvalanmış toplam 20 maddeye verdikleri cevaplardan elde edilen veriler kullanılmıştır. Bu veri setinde her bir madde takımında yer alan madde sayıları eşit olmadığı için $bx(m:t)$ desenine ilişkin dengelenmemiş bir veri seti ile çalışılmıştır. Maddelerin, madde takımlarına dağılımları Tablo 4.1'de verilmiştir.

Tablo 4.1: Maddelerin Madde Takımlarına Dağılımı

Madde Takımı Numarası	Madde Takımlarında Yer Alan Madde Numaraları	Madde Takımlarında Yer Alan Madde Sayıları
1	1, 2	2
2	3, 4, 5	3
3	6, 7, 8	3
4	9, 10, 11	3
5	12, 13, 14	3
6	15, 16, 17, 18, 19, 20	6
		Toplam: 20 Madde

Madde takımı etkisi ele alınmadan yürütülen G kuramı analizleri iki farklı şekilde gerçekleştirilmiştir:

1. Madde düzeyinde bireylerin (b) maddelerle (m) çaprazlandığı tesadüfi desen (bxm) : Çalışma kapsamında ele alınan 1500 öğrenci ve 20 maddeden oluşan veri seti kullanılmıştır.
2. Madde takımı düzeyinde bireylerin (b) madde takımlarıyla (t) çaprazlandığı tesadüfi desen (bxt) : Çalışma kapsamında ele alınan 1500 öğrenci ve 6 madde takımından oluşan veri seti kullanılmıştır. Her bir madde takımı puanı, madde takımlarını oluşturan maddelerin ortalaması alınarak elde edilmiştir.

Madde takımı etkisi ele alınarak yürütülen G kuramı analizlerinde ise 1500 öğrencinin altı madde takımında yuvalanmış 20 maddeye verdikleri cevapların çaprazlandığı tesadüfi desen [$bx(m:t)$] kullanılmıştır. Her bir madde takımında yer alan madde sayısı birbirinden farklı olduğu için veri dengelenmemiş yapıdadır ve analizler bu durum dikkate alınarak gerçekleştirilmiştir.

Yürütülen analizler sonucu elde edilen değerler Tablo 4.2’de verilmektedir.

Tablo 4.2: Güvenirlik Katsayıları

		<i>Madde takımı etkisinin ele alınmadığı durum</i>		<i>Madde takımı etkisinin ele alındığı durum [bx(m:t)]</i>
		<i>Madde düzeyi (bxm)</i>	<i>Madde takımı düzeyi (bxt)</i>	
Güvenirlik Katsayısı	G	0,776	0,762	0,761
	Phi	0,757	0,712	0,711

Tablo 4.2 incelendiğinde, çalışmada ele alınan 20 maddeden oluşan test verilerine madde düzeyinde uygulanan genellenebilirlik kuramı analizleri sonucu elde edilen G katsayısı 0,78 ve Phi katsayısı ise 0,76 olarak okunmaktadır. Her bir madde takımında yer alan maddelerin ortalaması alınarak elde edilen altı madde takımı için ise elde edilen G katsayısı 0,76 ve Phi katsayısı ise 0,71’dir. Bu durumda madde düzeyinde kestirilen G katsayısı, madde takımı düzeyinden yaklaşık 0,02 daha yüksek çıkmıştır. Madde düzeyinde kestirilen Phi katsayısı ise, madde takımı düzeyinden yaklaşık 0,07 daha yüksek çıkmıştır. Tablo 4.2’de okunan en düşük güvenilirlik değerleri ise madde takımı etkisi ele alındığı durumda kestirilmiştir.

Madde takımı düzeyinde ve madde takımı etkisi ele alınarak kestirilen G ve Phi katsayıları birbirine çok yakın çıkmıştır. Ancak madde takımı düzeyinde, testte altı madde varmış gibi analizler yürütülmüştür. Bir diğer ifadeyle, madde takımlarında yer alan maddelerin ortalamaları alınarak altı madde takımına ait puanlar üzerinden güvenilirlik analizleri yapılmıştır. Bunun yanı sıra, madde takımı etkisi dahil edildiğinde 20 maddelik veri üzerinden analizler yürütülmüştür. Madde sayıları bu kadar farklıken (6 ve 20 madde) elde edilen güvenilirlik değerlerinin birbirine çok yakın çıkması, kestirimlerden birinde olduğundan yüksek bir kestirim olduğu şeklinde yorumlanabilir. Sadece altı maddede, 20 maddelik veriye yakın kestirilen güvenilirlik değerlerinin, olduğundan yüksek kestirildiği düşünülebilir.

Sonuç olarak, Tablo 4.2'ye göre madde takımı etkisi ele alınmadığında kestirilen G ve Phi katsayılarının madde takımı etkisi ele alındığında kestirilen G ve Phi katsayılarından daha yüksek olduğu görülmektedir. Madde takımı etkisi ele alınmadığında güvenirliliğin olduğundan yüksek kestirilmesi alan yazındaki bazı çalışmaların bulguları ile uyumludur (Hendrickson, 2001; Lee ve Frisbie, 1999; Lee ve Frisbie, 1997; Wainer ve Thissen, 1996; Sireci, Thissen ve Wainer, 1991). Bu çalışmada da belirtilen çalışmalarla benzer bir sonuçla karşılaşılmıştır.

4.1.2. Alt Problem 1.2'ye İlişkin Bulgular ve Yorumlar

Alt problem 1.2. Maddelerin (m) madde takımlarında (t) yuvalandığı, birey (b) değişkeninin ise bunlarla çaprazlandığı dengelenmemiş $bx(m:t)$ deseninin genellenebilirlik kuramı sonuçları nasıldır?

Çalışmada maddelerin (m) madde takımlarında (t) yuvalandığı, birey (b) değişkeninin ise bunlarla çaprazlandığı dengelenmemiş $bx(m:t)$ desenine ilişkin yürütülen genellenebilirlik (G) ve karar (K) çalışmalarından elde edilen bulgular sırasıyla aşağıda verilmiştir.

4.1.2.1. Alt Problem 1.2.1'e İlişkin Bulgular ve Yorumlar

Alt problem 1.2.1. Dengelenmemiş desende G çalışması sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama oranları nasıldır?

Dengelenmemiş $bx(m:t)$ deseniyle yapılan G çalışması sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama oranları Tablo 4.3'te verilmiştir.

Tablo 4.3: Dengelenmemiş $bx(m:t)$ Desenine Ait G Çalışması Sonuçları

Varyans kaynağı	Kareler toplamı	Sd	Kareler ortalaması	Varyans değeri	Varyans yüzdesi
b	1305,54992	1499	0,87095	0,03314	13,13
t	403,12864	5	80,62573	0,01323	5,24
m:t	228,05199	14	16,28943	0,01073	4,25
bt	1540,90457	7495	0,20559	0,00457	1,81
bm:t,e	4003,61443	20986	0,19078	0,19078	75,57
Toplam	7481,24955	29999	98,18248	0,25245	100

Tablo 4.3'teki dengelenmemiş $bx(m:t)$ desenine ait G çalışması sonucu kestirilen varyans değerleri ve toplam varyansı açıklama oranları incelendiğinde, birey (b) ana etkisine ait varyansın 0,033'lük değeri ile toplam varyansın % 13,1'ini açıkladığı görülmektedir. Bu değer tabloda okunan en yüksek ikinci değerdir. Bu durum bireyler arası farklılığın ortaya çıkarılabildiğinin göstergesidir ve istenen bir durumdur.

Madde takımı (t) ana etkisi için kestirilen varyans bileşeni 0,013'lük değeri ile toplam varyansın % 5,2'sini açıklamaktadır. Bu değer küçük olması, madde takımlarının güçlü düzeyleri arasındaki farklılığın bir madde takımından diğerine çok fazla değişmediğinin göstergesidir.

Çalışmada, her bir madde takımında yer alan madde sayıları birbirinden farklıdır ve maddelerin madde takımlarında yuvalandığı $m:t$ etkisine ait varyans bileşeni 0,011'lik değeri ile toplam varyansın % 4,3'ünü açıklamaktadır. Maddelerin madde takımları içinde yuvalanması sonucunda $m:t$ etkisine ait varyans bileşeni ($\sigma^2_{m:t}$), madde (m) ana etkisi (σ^2_m) ve madde x madde takımı (mt) ortak etkisine (σ^2_{mt}) ait varyans bileşenlerinden oluşmaktadır (Brennan, 2001a). Bu değer küçük olması, madde takımlarında yer alan maddelerin güçlü düzeylerinin bireyler tarafından algılanmasında çok büyük bir farklılık olmadığı şeklinde yorumlanabilir.

Birey x madde takımı (bt) ortak etkisine ait varyans bileşeni 0,005'lik değeri ile toplam varyansın % 1,8'ini açıklamaktadır. Kestirilen bu varyans değerinin Tablo 4.3'te okunan en küçük varyans olduğu görülmektedir. Bu varyans değeri, birey-madde takımı etkileşiminden kaynaklı farklılıkların az olduğu şeklinde yorumlanabilir.

Artık etkisine ait varyans bileşeni ($bm:t,e$) 0,191'lik değeri ile toplam varyansın % 75,6'sını açıklamaktadır. Bu değer, toplam varyansı açıklama oranı bakımından Tablo 4.3'te okunan en yüksek değerdir. Yuvalanmış $bx(m:t)$ deseninde birey-madde (bm) ortak etkileşimi varyans değerinin ayrı olarak elde edilmesi mümkün değildir. Bu sebeple artık varyansın bir kısmında bu etkileşim oluşturmaktadır. Artık varyansın bu kadar yüksek olması birey-madde (bm) ortak etkileşiminin, birey-madde-madde takımı ortak etkileşiminin (bmt) veya diğer açıklanamayan değişkenlik kaynaklarının yol açabileceği etkilerin yüksek olabileceğini göstermektedir. Alan yazın tarandığında, dengelenmemiş desenlerin ele alındığı diğer çalışmalarda da artık varyansın oldukça yüksek kestirildiği görülmüştür (Nalbantoğlu Yılmaz, 2012; Hendrickson, 2001). Bu durum, desenin hem dengelenmemiş hem de yuvalanmış bir yapı göstermesinden kaynaklanıyor olabilir.

4.1.2.2. Alt Problem 1.2.2'ye İlişkin Bulgular ve Yorumlar

Alt problem 1.2.2. *Dengelenmemiş $bx(m:t)$ deseninde madde ve madde takımı sayılarının artırılıp azaltılması senaryolarına göre K çalışmasıyla kestirilen G ve Phi katsayıları sonuçları nasıldır?*

Her bir madde takımında farklı sayıda madde yer aldığından, dengelenmemiş $bx(m:t)$ deseninde madde ve her bir madde takımında yer alan madde sayılarının arttırılıp azaltılması ile yürütülen K çalışması sonucuna göre elde edilen G ve Phi katsayıları Tablo 4.4'te verilmiştir.

Tablo 4.4: Dengelenmemiş $bx(m:t)$ Desenine Ait K Çalışması Sonuçları

n_{m+}	n_t	$n_{m:t}$	\check{n}_t	G	Phi
16	3	4,5,7	2,84	0,710	0,637
16	6	2,2,2,3,3,4	5,57	0,722	0,677
16	6	2,2,3,3,3,3	5,82	0,723	0,679
16	8	2,2,2,2,2,2,2,2	8,00	0,726	0,691
20	3	5,7,8	2,89	0,749	0,671
20	6	2,3,3,3,3,6	5,26	0,761	0,711
20	6	2,3,3,4,4,4	5,71	0,762	0,715
20	8	2,2,2,2,3,3,3,3	7,69	0,766	0,728
40	3	12,13,15	2,97	0,840	0,750
40	6	4,4,6,6,9,11	5,22	0,854	0,797
40	6	3,5,7,7,9,9	5,44	0,855	0,800
40	8	2,2,4,4,6,6,8,8	6,67	0,859	0,811

n_{m+} : madde sayısı, n_t : madde takımı sayısı, $n_{m:t}$: madde takımlarında yer alan madde sayıları, $\check{n}_t: n_{m+}^2 / \sum_t n_{m:t}^2$

Tablo 4.4'te görüldüğü gibi madde takımlarında yer alan toplam madde sayısı ($n_{m+}=20$) sabit kalmak koşuluyla madde takımı sayısının azaltılması ($n_t=3$) sonucunda her bir madde takımında birbirinden farklı sayıda madde yer alması sonucu elde edilen G katsayısı 0,749 ve Phi katsayısı 0,671 olarak kestirilmiştir ($n_{m+}=20$, $n_t=3$, $\check{n}_t=2,89$). Madde takımlarında yer alan toplam madde sayısı ($n_{m+}=20$) sabit kalmak koşuluyla madde takımı sayısının artırılması ($n_t=8$) sonucunda ise her bir madde takımında birbirinden farklı sayıda madde yer alması sonucu elde edilen G katsayısı 0,766 ve Phi katsayısı 0,728 olarak kestirilmiştir ($n_{m+}=20$, $n_t=8$, $\check{n}_t=7,69$). Madde takımı sayısının azaltılması durumunda elde edilen G ve Phi katsayıları, araştırmada kullanılan verilerle elde edilen G katsayısına göre 0,011 ve Phi katsayısına göre ise 0,040 azalmaktadır. Madde takımı sayısının artırılması durumunda elde edilen G ve Phi katsayıları, araştırmada kullanılan verilerle elde edilen G katsayısına göre 0,0051 ve Phi katsayısına göre ise 0,017 artmaktadır. Sonuç olarak, madde sayısı sabit kalmak koşuluyla madde takımı sayılarını azaltmanın veya artırmanın G ve Phi katsayılarını çok fazla etkilemediği görülmektedir.

Madde takımı sayısı ($n_t=6$) sabit kalmak koşuluyla, her bir madde takımında yer alan madde sayısı ve buna bağlı olarak toplam madde sayısının azaltılması ($n_m=16$) sonucu elde edilen G katsayısı 0,722 ve Phi katsayısı ise 0,677 olarak kestirilmiştir ($n_{m+}=16$, $n_t=6$, $\check{n}_t=5,57$). Madde takımı sayısı ($n_t=6$) sabit kalmak koşuluyla, her bir madde takımında yer alan madde sayısı ve buna bağlı olarak toplam madde sayısının artırılması ($n_m=40$) sonucu elde edilen G katsayısı 0,855 ve Phi katsayısı ise 0,800 olarak kestirilmiştir ($n_{m+}=40$, $n_t=6$, $\check{n}_t=5,44$). Madde takımı sayısı sabit tutulduğunda madde sayısının azaltılması durumunda elde edilen G ve Phi katsayıları, araştırmada kullanılan verilerle elde edilen G katsayısına göre 0,039 ve Phi katsayısına göre ise 0,034 azalmaktadır. Madde sayısının artırılması durumunda elde edilen G ve Phi katsayıları, araştırmada kullanılan verilerle elde edilen G katsayısına göre 0,094 ve Phi katsayısına göre ise 0,089 artmaktadır. Sonuç olarak, madde takımı sayısı sabit kalmak koşuluyla madde takımlarında yer alan madde sayılarını azaltmanın veya artırmanın G ve Phi katsayılarını çok fazla etkilemediği görülmektedir.

Araştırmada kullanılan madde takımı ve madde takımlarında yer alan madde sayılarının birlikte artırılması durumunda ($n_{m+}=40$, $n_t=8$, $\check{n}_t=6,67$) G katsayısı 0,859 ve Phi katsayısı 0,811 olarak kestirilmiştir. Bu bulguya dayanarak, araştırmada kullanılan verilere göre madde takımı sayısı ile her bir madde takımında yer alan madde sayısının birlikte artırılması G katsayısında 0,098 ve Phi katsayısında 0,100 artırdığı görülmektedir. Bu durum, hem madde takımı hem de madde takımlarında yer alan madde sayılarının artırılmasının güvenilirlik katsayısını artırdığı sonucuna varılabilir.

Araştırmada kullanılan madde takımı sayısı ve toplam madde sayısı sabit kalmak şartıyla madde takımlarındaki madde sayısı farklılığının başlangıç durumuna göre azaltılması halinde ($n_{m+}=20$, $n_t=6$, $\check{n}_t=5,71$) G katsayısı 0,762 ve Phi katsayısı 0,715 olarak kestirilmiştir. Bu durumda madde takımı ve toplam madde sayısı sabit kalmak şartıyla madde takımlarında yer alan madde sayıları bakımından madde takımları arası değişkenliğin azaltılmasının G ve Phi katsayılarında artışa yol açtığı yorumu yapılabilir. Benzer durumlar madde takımı sayısının sabit olup toplam madde sayılarının 16 ve 40 olduğu durumlar için de geçerlidir. Madde sayısı 16

iken madde takımları arası deęişkenlięin az olduęu durumda ($n_{m+}=16$, $n_t=6$, $\check{n}_t=5,82$) kestirilen G ve Phi katsayıları deęişkenlięin fazla olduęu duruma ($n_{m+}=16$, $n_t=6$, $\check{n}_t=5,57$) göre daha yüksektir. Madde sayısı 40 iken madde takımları arası deęişkenlięin az olduęu durumda ($n_{m+}=40$, $n_t=6$, $\check{n}_t=5,44$) kestirilen G ve Phi katsayıları deęişkenlięin fazla olduęu duruma ($n_{m+}=40$, $n_t=6$, $\check{n}_t=5,22$) göre daha yüksektir.

4.1.2.3. Alt Problem 1.2.3'e İlişkin Bulgular ve Yorumlar

Alt problem 1.2.3. *Testte yer alan madde takımı sayısının ve her bir madde takımında yer alan madde sayısının G ve Phi katsayısına etkisi nasıldır?*

Dengelenmemiş $bx(m:t)$ deseni üzerinde yürütölen ve sonuçları Tablo 4.4'te verilen K çalıřmasından elde edilen bulgular doęrultusunda, hem madde takımı sayısının hem de madde takımlarında yer alan madde sayılarının artırılmasının güvenilirlięi olumlu şekilde etkiledięi görölmektedir. Ancak madde sayısı sabit tutulup testteki madde takımı sayısının artırılması sonucunda kestirilen G ve Phi katsayılarının, madde takımı sayısı sabit tutulup her bir madde takımında yer alan madde sayısının artırılması sonucu elde edilen G ve Phi katsayılarına göre daha fazla artışa yol açtıęı görölmektedir. Bu sonuç alan yazında da desteklenmektedir (Hendrickson, 2001; Lee ve Frisbie, 1999). Dięer bir ifadeyle, madde takımı sayısındaki artışın güvenilirlięe katkısı madde takımlarında yer alan madde sayısının güvenilirlięe katkısından daha fazladır.

Madde takımlarında gözlenen yerel madde baęımlılıęı, bir madde takımındaki madde sayısının artmasına baęlı olarak artış gösterebilir. Ancak madde takımlarının birbirlerinden baęımsız oldukları düşünölmektedir. Bu sebeple bir madde takımında yer alan madde sayısındaki artışın güvenilirlięe katkısı, madde takımı sayısının artışının güvenilirlięe katkısından daha düşük olması beklenen bir durumdur.

4.2. Alt Problem 2'ye İlişkin Bulgular ve Yorumlar

Alt problem 2. *Madde takımlarının değişen madde fonksiyonu üzerindeki etkisi nasıldır?*

Çalışma kapsamında ele alınan ve madde takımlarından oluşan maddeler üzerinde DMF belirleme çalışmaları yürütülmüştür. Çalışma kapsamında madde takımlarından oluşan testlerin DMF'ye etkisi belirlenmek amaçlandığı için madde takımı etkisini ele alan ve almayan iki yöntem kullanılmıştır. Kullanılan bu yöntemler Fukuhara ve Kamata (2011) tarafından önerilmiştir ve MTK'ya dayanmaktadır. Bu çalışmada DMF düzeylerinin belirlenmesinde her iki yöntem içinde 0,426 ölçütü B düzeyinde, 0,638 ölçütü ise C düzeyinde anlamlı DMF büyüklüğü olarak değerlendirilmiştir. WinBUGS yazılımı ile cinsiyet ve bölüme göre yürütülen DMF analiz sonuçları aşağıda ayrı başlıklar halinde verilmektedir.

4.2.1. Alt Problem 2.1'e İlişkin Bulgular ve Yorumlar

Alt problem 2.1. *Madde takımlarının yer aldığı testlerde, madde takımı etkisinin göz önünde bulundurulup bulundurulmaması durumlarında cinsiyete göre DMF gösteren madde sayısı ve kestirilen DMF düzeyi değişmekte midir?*

Çalışma kapsamında ele alınan test üzerinde cinsiyete göre yürütülen DMF analizlerinin madde takımı etkisinin ele alınıp alınmama durumlarına göre elde edilen bulgular Tablo 4.5'te verilmektedir.

Tablo 4.5: Ele Alınan İki Farklı Modelde Cinsiyete Göre Kestirilen DMF Büyüklükleri ve Düzeyleri

Madde	MTK DMF modeli (Madde takımı etkisi göz önünde bulundurulmadığında)			İki-faktörlü ÇBMTK'ya dayalı DMF modeli (Madde takımı etkisi göz önünde bulundurulduğunda)		
	DMF Büyüklüğü (β)	DMF Düzeyi	Avantajlı Grup	DMF büyüklüğü	DMF Düzeyi	Avantajlı Grup
1	0,0023			-0,01213		
2	0,0168			0,00307		
3	0,0235			0,00883		
4	0,41			0,36539		
5	0,0326			0,01159		
6	-0,6461 ***	C	Erkek	-0,79144 ***	C	Erkek
7	-0,4702 **	B	Erkek	-0,58357 **	B	Erkek
8	-0,3497 *	A	Erkek	-0,45286 **	B	Erkek
9	-0,0635			-0,08365		
10	0,7096 ***	C	Kadın	0,82231 ***	C	Kadın
11	-0,4955 **	B	Erkek	-0,76745 ***	C	Erkek
12	-0,0449			-0,07591		
13	-0,2459			-0,29884		
14	-0,5311 **	B	Erkek	-0,8571 ***	C	Erkek
15	-0,2145			-0,36132		
16	-0,2923 *	A	Erkek	-0,44404 **	B	Erkek
17	0,1441			0,22508		
18	0,17			0,25345		
19	-0,0386			-0,08705		
20	-0,2393			-0,31315		

* $|\beta| < 0,426$ (DMF yoktur ya da ihmal edilebilir düzeydedir.)

** $0,426 \leq |\beta| < 0,638$ (Orta düzeyde DMF vardır.)

*** $|\beta| \geq 0,638$ (Yüksek düzeyde DMF vardır.)

Tablo 4.5 incelendiğinde, MTK DMF modeli ile beş maddede (6, 7, 10, 11 ve 14) anlamlı düzeyde DMF belirlenmiştir. Bu beş maddede aynı zamanda madde takımları için iki faktörlü ÇBMTK'ya dayalı DMF model ile yürütülen analizlerde de anlamlı düzeyde DMF belirlenmiştir. Bunun yanında teste yer alan iki maddede (8 ve 16. maddeler) ise MTK DMF modelinde anlamlı düzeyde DMF kestirilmezken madde takımları için iki-faktörlü ÇBMTK'ya dayalı DMF modeli ile yürütülen analiz sonucunda anlamlı düzeyde DMF kestirilmiştir. Yani ÇBMTK'ya dayalı DMF modeli ile yürütülen analiz sonucunda anlamlı düzeyde DMF kestirilen madde sayısı MTK DMF modeli ile yürütülen analizden iki fazladır.

Cinsiyete göre yürütülen DMF analizlerinde kestirilen DMF düzeyleri incelendiğinde, kullanılan yöntemlerle elde edilen düzeyler arasında fark olduğu Tablo 4.5'te görülmektedir. 6, 7 ve 10. maddeler için kestirilen DMF düzeyleri her iki yöntemde de sırasıyla C, B ve C düzeylerinde kestirilmiştir. Ancak 8 ve

16.maddelerde MTK DMF modelinde A düzeyinde kestirilen DMF, ÇBMTK DMF modelinde B düzeyinde kestirilmiştir. Benzer şekilde 11 ve 14.maddelerde MTK DMF modelinde B düzeyinde kestirilen DMF, ÇBMTK DMF modelinde C düzeyinde kestirilmiştir.

ÇBMTK'ya dayalı DMF modeli ile yürütülen analiz sonucunda anlamlı düzeyde DMF kestirilen maddelerde DMF'nin kaynağını ortaya çıkarabilmek amacıyla, uzman kanısına başvurulmuştur. Buna göre;

- Erkekler lehine DMF içeren 6, 7 ve 8.maddelerin bağlı olduğu okuma parçası internet, sosyal medya ve duygular ile ilgilidir. Bu konuların günümüzde cinsiyete göre bir ayrımı olmadığı uzmanların büyük bir bölümü tarafından savunulmuş ve bu sebeple maddenin erkekler lehine bir yanlılık oluşturmadığı görüşü uzmanların neredeyse tamamı tarafından belirtilmiştir.
- Kadınlar lehine DMF içeren 10.madde ve erkekler lehine DMF içeren 11.maddenin bağlı olduğu okuma parçası makineler, endüstri, kimya ve mühendislik ile ilgili bir parçadır. Uzmanlar okuma parçasında yer alan konunun daha çok erkeklerin ilgisini çekebilecek türden olduğunu ifade etmektedirler. Bunun yanında 10.maddede sorulan ayrıntı, bir kumaş türü ile ilgili olduğundan bu maddenin kadınlar lehine yanlılık gösterebileceğine çalışmaya katılan uzmanlardan yedisi katılmaktadır. Diğer bir ifadeyle çalışmaya katılan uzmanların büyük çoğunluğu bu maddenin kadınlar lehine işlemesini beklediklerini bildirmişlerdir. Erkekler lehine DMF gösteren 11.madde ile ilgili uzman görüşleri incelendiğinde ise uzmanların yarısı madde takımı konusunun erkeklerin ilgisini çekebilecek türden olması gerekçesiyle erkekler lehine yanlı çıkabileceğini savunurken yarısı bu görüşe katılmamakta ve bu madde için cinsiyete göre yanlılığın olmadığını savunmuşlardır.
- Erkekler lehine DMF içeren 14.maddenin bağlı olduğu okuma parçası mekanik araçlar (feribot) ve bunların teknik aksaklıkları ile ilgili bir parçadır. Bu sebeple bu maddenin erkekler lehine yanlı çıkabileceği görüşü çalışmaya katılan uzmanların altısı tarafından desteklenmektedir. Diğer dört

uzman ise bu maddede cinsiyete göre bir yanlılık beklemediklerini belirtmişlerdir.

- Erkekler lehine DMF içeren 16.maddenin bağlı olduğu okuma parçası beyin göçü, ülke ekonomileri ve eğitim ile ilgili bir parçadır. Konunun cinsiyete göre erkekler lehine yanlılık oluşturmayacağı çalışmaya katılan uzmanların büyük çoğunluğu tarafından belirtilmiştir.

4.2.2. Alt Problem 2.2'ye İlişkin Bulgular ve Yorumlar

Alt problem 2.2. Madde takımlarının yer aldığı testlerde, madde takımı etkisinin göz önünde bulundurulup bulundurulmaması durumlarında bölüme göre DMF gösteren madde sayısı ve kestirilen DMF düzeyi değişmekte midir?

Madde takımı etkisini ele alan ve almayan yöntemler kullanılarak sayısal-eşit ağırlık bölümlerine göre yürütülen DMF analiz sonuçları Tablo 4.6'da verilmektedir.

Tablo 4.6: Ele Alınan İki Farklı Modelde Sayısal-Eşit Ağırlık Bölümlerine Göre Kestirilen DMF Büyüklükleri ve Düzeyleri

Madde	MTK DMF modeli (Madde takımı etkisi göz önünde bulundurulmadığında)			İki-faktörlü ÇBMTK'ya dayalı DMF modeli (Madde takımı etkisi göz önünde bulundurulduğunda)		
	DMF Büyüklüğü (β)	DMF Düzeyi	Avantajlı Grup	DMF büyüklüğü	DMF Düzeyi	Avantajlı Grup
1	0,2029			0,0023		
2	0,494 *	B	Eşit Ağırlık	0,436 **	B	Eşit Ağırlık
3	0,0554			0,0235		
4	0,378			0,41		
5	0,2941			0,0326		
6	-0,1443 *	A	Sayısal	-0,6461 ***	C	Sayısal
7	-0,162 *	A	Sayısal	-0,4702 **	B	Sayısal
8	0,0085			-0,3497		
9	-0,2619			-0,0635		
10	0,1385 *	A	Eşit Ağırlık	0,7096 ***	C	Eşit Ağırlık
11	-0,0237 *	A	Sayısal	-0,6955 ***	C	Sayısal
12	0,0942			-0,0449		
13	0,1545			-0,2459		
14	0,4074 *	A	Eşit Ağırlık	0,4689 **	B	Eşit Ağırlık
15	0,3739			-0,2145		
16	-0,1614			-0,2923		
17	0,0338			0,1441		
18	0,1382			0,17		
19	0,1025			-0,0386		
20	0,1618			-0,2393		

* $|\beta| < 0,426$ (DMF yoktur ya da ihmal edilebilir düzeydedir.)

** $0,426 \leq |\beta| < 0,638$ (Orta düzeyde DMF vardır.)

*** $|\beta| \geq 0,638$ (Yüksek düzeyde DMF vardır.)

Tablo 4.6 incelendiğinde, hem MTK DMF modeli ile hem de madde takımları için iki faktörlü ÇBMTK'ya dayalı DMF model ile yürütülen analizlerde ortak olarak sadece bir maddede (2.madde) anlamlı düzeyde DMF belirlenmiştir. Bunun yanı sıra beş maddede ise (6, 7, 10, 11 ve 14.maddeler) sadece madde takımları için iki faktörlü ÇBMTK'ya dayalı DMF modeli ile yürütülen analizlerde anlamlı düzeyde DMF belirlenmiştir.

Sayısal – eşit ağırlık bölümlerine göre yürütülen DMF analizlerinde kestirilen DMF düzeyi incelendiğinde kullanılan yöntemler arası farkın olduğu Tablo 4.6'da görülmektedir. 2.madde için kestirilen DMF düzeyleri her iki yöntemde de B düzeyinde kestirilmiştir. Ancak 6, 7, 10, 11 ve 14.maddelerde kestirilen DMF düzeyleri farklılaşmaktadır. 7 ve 14.maddelerde MTK DMF modelinde A düzeyinde kestirilen DMF, ÇBMTK DMF modelinde B düzeyinde kestirilmiştir. Yani MTK DMF modelinde önemsenmeyecek düzeyde çıkan DMF anlamlı düzeyde kestirilmiştir. Benzer şekilde 6, 10 ve 11.maddelerde MTK DMF modelinde A düzeyinde kestirilen DMF, ÇBMTK DMF modelinde C düzeyinde kestirilmiştir.

ÇBMTK'ya dayalı DMF modeli ile yürütülen analiz sonucunda anlamlı düzeyde DMF kestirilen maddelerde DMF'nin kaynağını ortaya çıkarabilmek amacıyla, uzman kanısına başvurulmuştur. Buna göre;

- Eşit ağırlık bölümü öğrencilerinin lehine DMF içeren 2.maddenin bağlı olduğu okuma parçası uzay, uzay modülleri ve astronotlar ile ilgilidir. Bu konuların bölüme göre bir ayrımı olmadığı uzmanların büyük bir bölümü tarafından savunulmuş ve bu sebeple maddenin eşit ağırlık bölümü öğrencilerinin lehine bir yanlılık oluşturmadığı görüşü uzmanların büyük bir kısmı tarafından belirtilmiştir.
- Sayısal bölüm öğrencilerinin lehine DMF içeren 6 ve 7.maddelerin bağlı olduğu okuma parçası internet, sosyal medya ve duygular ile ilgilidir. Araştırma kapsamında görüşleri alınan uzmanların büyük çoğunluğu 6.maddenin sayısal bölüm öğrenciler lehine yanlı çıkmasını beklemediklerini belirtmektedirler. Hatta bu uzmanlardan ikisi bölüm karşılaştırması yapılırsa bu maddenin psikolojik bir tema içerdiğinden dolayı

sayısal yerine eşit ağırlık öğrencilerinin lehine çıkmasını beklediklerini ifade etmişlerdir. 7.maddeye yönelik uzman görüşleri incelendiğinde uzmanların çoğunluğu bu maddede de bölüme göre yanlılık beklemediklerini belirtmişlerdir.

- Aynı madde takımında yer alan ve eşit ağırlık bölümü öğrencileri lehine DMF içeren 10.madde ile sayısal bölüm öğrencileri lehine DMF içeren 11.maddenin bağlı olduğu okuma parçası makineler, endüstri, kimya ve mühendislik ile ilgili bir parçadır. Uzman görüşleri incelendiğinde 10.maddenin eşit ağırlık bölümü öğrencileri lehine işlemesine uzmanların neredeyse tamamı katılmamaktadır. 11.madde ile ilgili ise uzmanlar görüş ayrılığına düşmüşlerdir. Beş uzman sayısal bölüm öğrencilerinin fen konularına daha aşina olmaları ve daha fazla bilgi sahibi olmaları sebebiyle bu maddede sayısal öğrenciler lehine yanlılık olabileceğini düşünmektedirler. Diğer beş uzman ise 10.maddede olduğu gibi bu maddenin herhangi bir bölüm lehine yanlılık teşkil etmediğini düşünmektedirler.
- Eşit ağırlık bölümü öğrencilerinin lehine DMF içeren 14.maddenin bağlı olduğu okuma parçası mekanik araçlar (feribot) ve bunların teknik aksaklıkları ile ilgili bir parçadır. Bu madde ile görüş alınan uzmanların büyük çoğunluğu bir yanlılık olmadığını belirtmişlerdir. Ancak üç uzman konunun uluslararası ticaret, uluslararası işletme, işletme, ekonomi vb. bölümlerde okuyan öğrencilerin ilgi alanları ile bağlantılı olduğundan ve bu bölümlere de eşit ağırlık puanı ile öğrenci alındığından, bu öğrencilerin aşina olması sebebiyle maddenin eşit ağırlık bölümü öğrencilerinin lehine olabileceğini belirtmişlerdir.

Madde takımı etkisini ele alan ve almayan yöntemler kullanılarak sayısal-sözel bölümlerine göre yürütülen DMF analiz sonuçları Tablo 4.7’de verilmektedir.

Tablo 4.7: Ele Alınan İki Farklı Modelde Sayısal-Sözel Bölümlerine Göre Kestirilen DMF Büyüklükleri ve Düzeyleri

Madde	MTK DMF modeli (Madde takımı etkisi göz önünde bulundurulmadığında)			İki-faktörlü ÇBMTK’ya dayalı DMF modeli (Madde takımı etkisi göz önünde bulundurulduğunda)		
	DMF Büyüklüğü (β)	DMF Düzeyi	Avantajlı Grup	DMF büyüklüğü	DMF Düzeyi	Avantajlı Grup
1	0,15716			0,1324		
2	0,57846 **	B	Sözel	0,5143 **	B	Sözel
3	0,0693			0,0376		
4	0,38313			0,3281		
5	0,22663			0,1749		
6	-0,01287			0,0211		
7	0,0853			0,1708		
8	-0,08582			-0,0842		
9	-0,19309			-0,1924		
10	0,0484			0,0566		
11	-0,11013			-0,1126		
12	0,01842			-0,0064		
13	0,15176			0,1518		
14	0,34438			0,3189		
15	0,21164			0,2474		
16	-0,27158			-0,286		
17	-0,10152			-0,0337		
18	0,09339			0,203		
19	-0,01535			0,02		
20	-0,16807			-0,1486		

** $0,426 \leq |\beta| < 0,638$ (Orta düzeyde DMF vardır.)

Tablo 4.7 incelendiğinde, hem MTK DMF modeli ile hem de madde takımları için iki faktörlü ÇBMTK’ya dayalı DMF model ile yürütülen analizlerde de sadece bir maddede (2.madde) anlamlı ve B düzeyinde DMF belirlenmiştir. Diğer bir ifade ile sayısal – sözel bölümlerine göre yürütülen DMF analizleri sonucunda kullanılan yöntemler arasında DMF gösteren madde sayısı ve düzeyleri arasında bir farklılık ortaya çıkmamıştır.

ÇBMTK’ya dayalı DMF modeli ile yürütülen analiz sonucunda anlamlı düzeyde DMF kestirilen maddede DMF’nin kaynağını ortaya çıkarabilmek amacıyla, uzman kanısına başvurulmuştur. Buna göre;

- Sözel bölüm öğrencilerinin lehine DMF içeren 2.maddenin bağlı olduğu okuma parçası uzay, uzay modülleri ve astronomlar ile ilgilidir. Bu konuların

bölüme göre bir ayrımı olmadığı uzmanların büyük bir bölümü tarafından savunulmuş ve bu sebeple maddenin sözel bölüm öğrencilerinin lehine bir yanlılık oluşturmadığı görüşü uzmanların büyük bir kısmı tarafından belirtilmiştir.

Madde takımı etkisini ele alan ve almayan yöntemler kullanılarak eşit ağırlık-sözel bölümlerine göre yürütülen DMF analiz sonuçları Tablo 4.8'de verilmektedir.

Tablo 4.8: Ele Alınan İki Farklı Modelde Eşit Ağırlık-Sözel Bölümlerine Göre Kestirilen DMF Büyüklükleri ve Düzeyleri

Madde	MTK DMF modeli (Madde takımı etkisi göz önünde bulundurulmadığında)			İki-faktörlü ÇBMTK'ya dayalı DMF modeli (Madde takımı etkisi göz önünde bulundurulduğunda)		
	DMF Büyüklüğü (β)	DMF Düzeyi	Avantajlı Grup	DMF büyüklüğü	DMF Düzeyi	Avantajlı Grup
1	0,2131			0,2049		
2	0,4881 **	B	Sözel	0,4974 **	B	Sözel
3	0,1202			0,0549		
4	0,3659			0,3802		
5	0,3149			0,296		
6	0,1172			0,1418		
7	0,0983			0,159		
8	-0,0317			0,0062		
9	-0,254			-0,2637		
10	0,1269			0,1377		
11	0,0283			0,0216		
12	0,113			0,093		
13	0,1786			0,1541		
14	0,3477			0,4035		
15	0,2791			0,3727		
16	-0,1522			-0,1605		
17	-0,0193			0,0372		
18	0,0626			0,1393		
19	0,0571			0,1078		
20	0,1301			0,1602		

**0,426 \leq |\mathbf{\beta}|< 0,638 (Orta düzeyde DMF vardır.)

Tablo 4.8 incelendiğinde, hem MTK DMF modeli ile hem de madde takımları için iki faktörlü ÇBMTK'ya dayalı DMF model ile yürütülen analizlerde de sadece bir maddede (2.madde) anlamlı B düzeyinde DMF belirlenmiştir. Diğer bir ifade ile eşit ağırlık – sözel bölümlerine göre yürütülen DMF analizleri sonucunda kullanılan

yöntemler arasında DMF gösteren madde sayısı ve düzeyleri arasında bir farklılık ortaya çıkmamıştır.

ÇBMTK'ya dayalı DMF modeli ile yürütülen analiz sonucunda anlamlı düzeyde DMF kestirilen maddede DMF'nin kaynağını ortaya çıkarabilmek amacıyla, uzman kanısına başvurulmuştur. Buna göre;

- Sözel bölüm öğrencilerinin lehine DMF içeren 2.maddenin bağlı olduğu okuma parçası uzay, uzay modülleri ve astronotlar ile ilgilidir. Bu konuların bölüme göre bir ayrımı olmadığı uzmanların büyük bir bölümü tarafından savunulmuş ve bu sebeple maddenin sözel bölüm öğrencilerinin lehine bir yanlılık oluşturmadığı görüşü uzmanların büyük bir kısmı tarafından belirtilmiştir.

5. SONUÇ ve ÖNERİLER

Bu bölümde araştırmmanın bulgu ve yorumlarına dayalı olarak ulaşılan sonuçların özetine ve bu sonuçlardan yola çıkarak geliştirilen önerilere yer verilmiştir.

5.1. Sonuçlar

Araştırmadan elde edilen bulgulara dayalı olarak ulaşılan sonuçlar, alt problemlerin sırasına uygun olarak aşağıda açıklanmıştır.

5.1.1. Alt Problem 1'e İlişkin Sonuçlar

1. Öncelikle madde takımlarından oluşan testlerin güvenilirliği kestirilken G kuramından faydalanılmıştır. Madde takımlarından oluşan testlerin güvenilirliği hem madde takımı etkisinin göz önünde bulundurulmadığı madde ve madde takımları düzeylerinde hem de madde takımı etkisinin dahil edildiği yuvalanmış desen için kestirilmiştir. Kestirilen bu değerler içinde madde düzeyinde elde edilen güvenilirlik katsayısı diğer katsayılardan daha yüksek çıkmıştır. G kuramında tek yüzeyli desenlerde kestirilen G katsayısı Cronbach alfa katsayısına eşit çıkmaktadır (Brennan, 2001a). Sonuç olarak madde takımlarından oluşan testlerde madde takımı etkisi göz önünde bulundurulmadığında madde düzeyinde kestirilen ve Cronbach alfa katsayısına da eşit çıkan güvenilirlik katsayısı ile olduğundan yüksek, bir diğer ifade ile yanlı bir sonuç elde edilmektedir. Madde takımı etkisi dahil edilmediğinde kestirilen güvenilirlik değerinin testi oluşturan madde sayısını iki katına çıkarıldığında elde edebilecek bir güvenilirliğe ulaşabilecek kadar yüksek kestirildiği durumlar ile karşılaşmak dahi mümkündür (Wainer, Sireci ve Thissen, 1991).
2. Test geliştirme sürecinde pek çok avantajı sebebiyle madde takımlarının testlerde kullanılması sıklıkla karşılaşılan bir durumdur. Araştırma kapsamında elde edilen sonuçlar doğrultusunda güvenilirliği artırmak adına madde takımı sayısındaki artışın, madde takımlarını oluşturan madde sayısındaki artıştan daha etkili olduğu sonucuna ulaşılmıştır. Bu durum G kuramı çerçevesinde yürütülen G çalışmaları ile desteklenmektedir (Bkz.

Tablo 4.3). Madde takımlarının toplam varyansa etkisi maddelerin madde takımlarında yuvalanmasıyla elde edilen varyans değerinden daha büyüktür.

3. Bir testte yer alan madde takımlarını oluşturan madde sayılarındaki çeşitlilik test sonuçlarından elde edilebilecek güvenilirliği etkilemektedir. Araştırmada testte yer alan madde takımlarını oluşturan madde sayıları bir madde takımından diğerine fazlasıyla değişmekte ise bu durumun güvenilirliği olumsuz etkilediği sonucuna ulaşılmıştır.

5.1.2. Alt Problem 2'ye İlişkin Sonuçlar

1. Bu çalışma kapsamında gerçek veri üzerinde yürütülen DMF çalışmaları sonucunda madde takımı etkisini ele alan DMF belirleme yöntemi ile anlamlı düzeyde DMF gösterdiği belirlenen madde sayısı madde takımı etkisini ele almayan yöntemle oranla daha fazladır.
2. Madde takımlarının yerel madde bağımsızlığı varsayımını ihlal etmelerinden dolayı tek boyutlu MTK'ya dayanan yöntemler ile kestirilen DMF büyüklükleri olduğundan küçük kestirilmektedir. Ayrıca alan yazında yürütülen DMF çalışmaları incelendiğinde, madde takımı etkisi ele alınmadan yürütülen DMF analizleri sonucunda madde düzeyinde önemsenmeyecek düzeyde çıkan DMF'nin madde takımı etkisi dahil edildiğinde belirginleştiği sonucuna varılmıştır (Fukuhara ve Kamata, 2011; Fukuhara, 2009; Sedivy, 2009; Wainer, Sireci ve Thissen, 1991). Bu durum veri sayısı yeterince fazla olduğunda çok ciddi farklılıklara yol açmamakta ancak örneklem ve DMF küçüldükçe madde takımı etkisini göz önünde bulunduran yöntemlerin daha iyi bir performans sergilediği görülmektedir.

Sonuç olarak test geliştirme çalışmalarında madde takımlarının kullanım sıklığının artması bu araçların doğru istatistiksel yollar ile analiz edilmesini bu testlerden elde edilecek sonuçların doğru yorumlanabilmesi açısından bir zorunluluk haline getirmektedir.

5.2. Öneriler

Araştırma bulgularına dayalı olarak geliştirilen ve ileride yapılabilecek başka araştırmalarda değerlendirilebilecek öneriler aşağıda verilmiştir.

5.2.1. Araştırma Sonuçlarından Çıkan Öneriler

1. Madde takımlarından oluşan testlerin güvenilirliği kestirilirken yerel madde bağımlılığını göz önünde bulunduran kestirim yöntemleri daha yansız kestirime ulaşmak adına tercih edilebilir.
2. Madde takımlarından oluşan testlerde daha güvenilir ölçme sonuçları elde edebilmek adına testlerde yer alan madde takımlarındaki madde sayısı yerine madde takımı sayısı artışına gitmek daha etkili olduğundan bu seçenek tercih edilebilir. Diğer bir ifadeyle ortak bir uyarana bağlı olarak hazırlanan madde takımlarının sayısını artırmak güvenilirliği artırmada etkili olduğu için bu yol izlenebilir.
3. Madde takımlarından oluşan testlerin güvenilirliğini artırmak adına testte yer alan madde takımlarını oluşturan madde sayılarının birbirine yakın olması bir diğer ifadeyle bir madde takımından diğerine çok fazla değişmemesi sağlanabilir.

5.2.2. Araştırmacılara Dönük Öneriler

1. Madde takımlarından oluşan testlerde DMF büyüklüğü kestirmede kullanılabilen madde takımları için iki faktörlü ÇBMTK modeline dayanan yöntem, başka test durumlarında da kullanılabilir. Eğitimdeki pek çok durumun yerel madde bağımlılığına yol açma potansiyeli bulunmaktadır (Yen, 1993). Örneğin, standart bir fen testi fizik, kimya, biyoloji gibi alt temalardan oluşmaktadır. Bu tür bir testte bu alt temalar kümelenerek madde takımı gibi düşünülebilir ve DMF belirlenirken bu çalışma kapsamında kullanılan yöntem ile analizler gerçekleştirilebilir.
2. Bu araştırmanın tek bir veri seti üzerinden ve bu sebeple görece az sayıda madde üzerinde gerçekleştirilmesi bir sınırlılık olarak görülebilir. Ancak madde takımlarının güvenilirliğe (Hendrickson, 2001; Lee ve Frisbie, 1999; Lee ve Frisbie, 1997; Wainer, 1995; Sireci, Thissen ve Wainer, 1991) ve

değişen madde fonksiyonuna etkisi (Fukuhara ve Kamata, 2011; Wainer, 1995) üzerine yapılan araştırmalarda da benzer durumlarla karşılaşılmaktadır. Bununla birlikte hem farklı veri setleri (sadece madde takımlarından oluşan veya bağımsız maddeler ve madde takımlarının bir arada yer aldığı) üzerinde hem de daha çok maddeden oluşan veri setleri üzerinde araştırma yapmak, daha güvenilir sonuçlara ulaşmak adına faydalı olacaktır.

3. Bu araştırmada gerçek veri kullanılarak analizler gerçekleştirilmiştir. Madde takımı etkisini farklı ölçme durumlarda daha detaylı değerlendirebilmek adına simülasyon çalışmaları yürütülebilir.

Sonuç olarak, madde takımlarından oluşan testler için bu çalışma kapsamında da ele alınan güvenilirlik ve geçerlik yöntemleri gibi madde takımlarının yerel madde bağımlılığı gösterme durumlarını görmezden gelmeyen yöntemler ile daha yansız sonuçlar elde edilmesi hedeflenmelidir.

KAYNAKÇA

- Ackerman, T. A. (1987). *The Robustness of LOGIST and BILOG IRT Estimation Programs to Violations of Local Independence*. Paper Presented at the Annual Meeting of the American Educational Research Association, Washington, DC.
- Angoff, W. H. (1993). Perspectives on differential item functioning methodology. W. Holland ve H. Wainer içinde, *Differential item functioning* (s. 3-24). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Baykul, Y. (2000). *Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması*. Ankara: ÖSYM.
- Bell, J. F. (1985). Generalizability Theory: The Software Problem. *Journal of Educational and Behavioral Statistics*, 10 (1), 19-29.
- Bloch, R. ve Norman, G. (2011). *G String 4 User Manual (Version 6.1.1)*. Hamilton, Ontario, Canada: Web: http://fhspcrd.mcmaster.ca/g_string/download/g_string_4_manual_611.pdf adresinden 10 Mayıs 2013'te alınmıştır.
- Bradlow, E. T., Wainer, H. ve Wang, X. (1999). A Bayesian Random Effects Model for Testlets. *Psychometrika*, 64, 153-168.
- Brennan, R. L. (2001a). *Generalizability Theory*. New-York: Springer-Verlag.
- Brennan, R. L. (2001b). *Manual for urGENOVA (Version 2.1)*. Iowa City: Iowa Testing Programs Occasional Papers. Web: <http://www2.education.uiowa.edu/centers/casma/computer-programs.aspx#genova> adresinden 10 Mayıs 2013'te alınmıştır.
- Clauser, B. E. ve Mazor, K. M. (1998). Using Statistical Procedures to Identify Differentially Functioning Test Items. *Educational Measurement: Issues and Practice*, 17 (1), 31-47.
- DeMars, C. E. (2006). Application of the Bi-Factor Multidimensional Item Response Theory Model to Testlet-Based Tests. *Journal of Educational Measurement*, 43(2), 145-168.
- Dorans, N. J. ve Holland, P. W. (1992). *DIF Detection and Description: Mantel-Haenszel and Standardization*. Princeton, N. J.: Educational Testing Service.
- Embretson, S. E. ve Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Fukuhara, H. (2009). *A Differential Item Functioning Model for Testlet-Based Items Using A Bi-Factor Multidimensional Item Response Theory Model: A Bayesian Approach*. Yayınlanmamış Doktora Tezi: Florida State University College of Education.
- Fukuhara, H. ve Kamata, A. (2007). *DIF Detection in a Presence of Locally Dependent Items*. Paper presented at the annual meeting of the Florida Educational Research Association: Tampa, FL.

- Fukuhara, H. ve Kamata, A. (2011). A Bifactor Multidimensional Item Response Theory Model for Differential Item Functioning Analysis on Testlet-Based Items. *Applied Psychological Measurement*, 35 (8), 604-622.
- Gessaroli, M. E. ve Folske, J. C. (2002). Generalizing the Reliability of Tests Comprised of Testlets. *International Journal of Testing*, 2(3-4), 277-295.
- Glas, C. A., Wainer, H. ve Bradlow, E. T. (2002). MML and EAP Estimation in Testlet-Based Adaptive Testing. W. J. van der Linden ve C. A. Glas içinde, *Computerized Adaptive Testing: Theory and Practice* (s. 271-287). Dordrecht: Kluwer Academic Publishers.
- Güler, N. (2008). *Klasik Test Kuramı, Genellenabilirlik Kuramı ve Rasch Modeli Üzerine Bir Araştırma*. Yayınlanmamış Doktora Tezi, Hacettepe Üniversitesi, Ankara.
- Güler, N., Kaya Uyanık, G. ve Taşdelen Teker, G. (2012). *Genellenebilirlik Kuramı*. Ankara: Pegem Akademi Yayıncılık.
- Hambleton, R. K. ve Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. USA: Kluwer-Nijhoff Publishing.
- Hambleton, R. K., Swaminathan, H. ve Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. California: SAGE Publications.
- Hendrickson, A. B. (2001). *Reliability of Scores from Tests Composed of Testlets: A Comparison of Methods*. Seattle, WA, April 11-13: Paper presented at the Annual Meeting of the National Council on Measurement in Education.
- Jarjoura, D. ve Brennan, R. L. (1981). *Three Variance Components Models for Some Measurement Procedures in Which Unequal Numbers of Items Fall into Discrete Categories*. Iowa City: ACT Technical Bulletin, Number 37.
- Lawrence, I. M. (1995). *Estimating reliability for tests composed of item sets* (Research Rep. No. 95-18). Princeton, NJ: Educational Testing Service.
- Lee, G. ve Frisbie, D. A. (1997). *A Generalizability Approach to Evaluating the Reliability of Testlet-Based Test Scores*. Chicago: Paper presented at the Annual Meeting of the National Council on Measurement in Education.
- Lee, G. ve Frisbie, D. A. (1999). Estimating Reliability Under a Generalizability Theory Model for Test Scores Composed of Testlets. *Applied Measurement in Education*, 12(3), 237-255.
- Lee, G., Brennan, R. L. ve Frisbie, D. A. (2000). Incorporating the Testlet Concept in Test Score Analyses. *Educational Measurement: Issues and Practice*, 19 (4), 9-15.
- Lee, G., Dunbar, S. B. ve Frisbie, D. A. (2001). The Relative Appropriateness of Eight Measurement Models for Analyzing Scores from Test Composed of Testlets. *Educational and Psychological Measurement*, 61(6), 958-975.
- Lee, G., Kolen, M. J., Frisbie, D. A. ve Ankenmann, R. D. (2001). Comparison of Dichotomous and Polytomous Item Response Models in Equating Scores from Test Composed of Testlets. *Applied Psychological Measurement*, 25, 357-372.

- Lee, G. ve Park, I. (2012). A Comparison of The Approaches of Generalizability Theory and Item Response Theory in Estimating The Reliability of Test Scores for Testlet-Composed Tests. *Asia Pasific Education Review*, 13(1), 47-54.
- Li, Y., Bolt, D. M. ve Fu, J. (2006). A Comparison of Alternative Models for Testlets. *Applied Psychological Measurement*, 30(1), 3-21.
- Masters, G. N. (1988). Item Discrimination: When more is worse. *Journal of Educational Measurement*, 25(1), 15-29.
- Millsap, R. E. ve Everson, H. T. (1993). Methodology Review: Statistical Approaches for Assessing Measurement Bias. *Applied Psychological Measurement*, 17 (4), 297-334.
- Nalbantođlu Yılmaz, F. (2012). *Genellenebilirlik Kuramında Dengelenmiş ve Dengelenmemiş Desenlerin Karşılaştırılması*. Yayınlanmamış Doktora Tezi: Ankara Üniversitesi.
- Nalbantođlu Yılmaz, F. ve Uzun Başusta, B. (2012). Genellenebilirlik Kuramıyla Dikiş Atma ve Alma Becerileri İstasyonu Güvenirliđinin Deđerlendirilmesi. *III. Ulusal Eđitimde ve Psikolojide Ölçme ve Deđerlendirme Kongresi*, (s. Sözlü Bildiri). Bolu.
- Narayanan, P. ve Swaminathan, H. (1996). Identification of Items That Show Nonuniform DIF. *Applied Psychological Measurement*, 20, 257-274.
- Özçelik, D. A. (2010). *Test Hazırlama Kılavuzu*. Ankara: Pegem Akademi.
- Sedivy, S. K. (2009). *Using Traditional Methods to Detect Differential Item Functioning in Testlet Data*. Yayınlanmamış Doktora Tezi: University of Wisconsin-Milwaukee.
- Shavelson, R. J. ve Webb, N. M. (1991). *Generalizability Theory: A Primer*. USA: SAGE Publications.
- Sireci, S. G., Thissen, D. ve Wainer, H. (1991). On the Reliability of Testlet-Based Tests. *Journal of Educational Measurement*, 28, 237-247.
- Spiegelhalter, D., Thomas, A. ve Best, N. (2003). *WinBUGS 1.4 [bilgisayar programı]*. Cambridge, UK: MRC Biostatistics Unit, Institute of Public Health.
- Thissen, D., Steinberg, L. ve Mooney, J. A. (1989). Trace Lines for Testlets: A Use of Multiple-Categorical-Response Models. *Journal of Educational Measurement*, 26(3), 247-260.
- Thompson, B. (2003). *Score Reliability*. Thousand Oaks, Calif.: SAGE Publications.
- Turgut, F. M. ve Baykul, Y. (2012). *Eđitimde Ölçme ve Deđerlendirme*. Ankara: Pegem Akademi.
- Vaughn, B. K. (2006). *A Hierarchical Generalized Linear Model of Random Differential Item Functioning for Polytomous Items: A Bayesian Multilevel Approach*. Yayınlanmamış Doktora Tezi: The Florida State University College of Education.
- Wainer, H. (1995). Precision and Differential Item Functioning on a Testlet-Based Test: The 1991 Law School Admissions Test as an Example. *Applied Measurement in Education*, 8, 157-186.

- Wainer, H. ve Kiely, G. L. (1987). Item Clusters and Computerized Adaptive Testing: A Case for Testlets. *Journal of Educational Measurement*, 24 (3), 185-201.
- Wainer, H. ve Lewis, C. (1990). Toward a Psychometrics for Testlets. *Journal of Educational Measurement*, 27(1), 1-14.
- Wainer, H. ve Thissen, D. (1996). How is reliability related to the quality of test scores? What is the effect of local dependence on reliability? *Educational Measurement: Issues and Practice*, 15(1), 22-29.
- Wainer, H. ve Wang, C. (2000). Using A New Statistical Model for Testlets to Score TOEFL. *Journal of Educational Measurement*, 37, 203-220.
- Wainer, H., Bradlow, E. T. ve Du, Z. (2000). Testlet response theory: An analog for the 3PL model useful in testlet-based adaptive testing. W. J. van der Linden ve C. A. Glas içinde, *Computerized Adaptive Testing: Theory and Practice* (s. 245-269). Dordrecht: Kluwer Academic Publishers.
- Wainer, H., Sireci, S. G. ve Thissen, D. (1991). Differential Testlet Functioning: Definitions and Detection. *Journal of Educational Measurement*, 28 (3), 197-219.
- Wang, W. C. ve Wilson, M. (2005). Assessment of Differential Item Functioning in Testlet-Based Items Using the Rasch Testlet Model. *Educational and Psychological Measurement*, 65 (4), 549-576.
- Wang, X., Bradlow, E. T. ve Wainer, H. (2002). A General Bayesian Model for Testlets: Theory and Application. *Applied Psychological Measurement*, 26(1), 109-128.
- Yen, W. M. (1993). Scaling Performance Assessment: Strategies for Managing Local Item Dependence. *Journal of Educational Measurement*, 30, 187-213

EKLER DİZİNİ

EK-1: Çalışma Kapsamında Ele Alınan İngilizce Yeterlik Sınavı Soruları

SECTION IV: TEXT COMPREHENSION

Questions 1-20

On your answer sheet, mark the alternative which best answers the question or completes the statement about the text.

(1) Last week, a pair of Russian space modules arrived on the Isle of Man, in the British Isles, home to the private space company Excalibur. The firm will work on updating the modules, and plans eventually to use them to provide extra room and supplies for the tourists and researchers it hopes to take into space. "They are not the nice shiny stuff, but they have been proven to be a reliable means of space travel," says former astronaut Leroy Chiao, who now works for Excalibur.

(2) The modules, which have never been launched, were built in the 1970s as part of Almaz, a military programme that sent cosmonauts into orbit to take photographs of the Earth. Excalibur has also bought four Almaz spacecraft, which could be launched sooner than the modules, to start ferrying NASA astronauts into space.

Adapted from: *New Scientist*, January 2011

1. The two space modules Excalibur has bought _____.

- a) were sent into space in the 1970s
- b) were originally designed for space tourism
- c) are ready to be launched
- d) have been tested and are safe

2. Excalibur _____.

- a) employs a former astronaut
- b) has been carrying NASA astronauts into space
- c) was actively involved in Almaz
- d) has recently launched four spacecraft into orbit

(1) Today, tough, rubbery plastics are found in thousands of consumer goods, ranging from common household items such as bags and storage bins to inner tyres and expensive medical equipment. But the materials used are highly vulnerable to damage caused by scratches, cuts and punctures. The result: landfills are full of plastic objects discarded because they broke, cracked or leaked, sometimes causing safety hazards. A new kind of plastic that scientists unveiled last week seems to be the solution because it can repair itself.

(2) Traditional approaches to repairing plastic materials involve heating damaged areas or applying patches. Using chemical fillings is also a common practice. The new material, on the other hand, requires something far simpler. When the molecules in the material are exposed to light, they disassemble and flow into the crack and the material gets healed. The technique has advantages, such as pinpoint targeting of the damaged area, and repairing objects that are still carrying a stress load.

(3) Smart materials with an in-built ability to repair damage caused by normal wear-and-tear could prove useful in transportation, construction, packaging and many other applications, the researchers said. Healable plastics offer an alternative to the damage-and-discard cycle, but several hurdles, including cost-ineffectiveness, remain before the innovation can be translated into industrial-scale production.

Adapted from: www.iafrica.com

3. In paragraph 1, which one is not mentioned about plastics?

- a) They are found in waste sites in large quantities.
- b) They are prone to be damaged easily.
- c) They are used less commonly in expensive equipment.
- d) They may be dangerous to the environment.

4. What does the new plastic rely on for repair?

- a) Patches
- b) Light
- c) Chemicals
- d) Heat

5. As we understand from the text, the new plastic material _____.

- a) can be produced at a lower cost than others
- b) does not break even under heavy load
- c) could extend the lifetime of products
- d) is already widely used in industry

(1) Social scientists have long known that social networks tend to have an "assortative" character—people are more likely to have friends who are similar in age, nationality or race, and less likely to be linked to people who are different. Psychologist Johan Bollen of the University of Indiana and colleagues wondered if people's moods is also a factor which draws them together on an online network like Twitter.

(2) To find out, they tracked 102,000 Twitter users over six months, analysing their 129 million, 140-letter-or-less tweets, using standard techniques from psychology. Specifically, they measured the emotional content of the tweets as reflected by the presence of positive or negative words. From this they could assess the "subjective well-being", or happiness, of the users through their tweets. The researchers indeed found that happier people—those using positive statements—tended to be tweeting and receiving tweets from people who were also happier. Similarly, less happy people tended to group together. According to Bollen, Twitter users are preferentially linked to those with whom they share a similar level of general happiness.

(3) Bollen admits they don't yet know why this is true. Happy or unhappy people may simply seek one another out, drawn by tweets expressing emotions similar to their own. Or, Bollen suggests, it could be that the emotions expressed, even in short tweets, have an infectious quality, lifting peoples' spirits or filling them with gloom, depending on what they read.

Adapted from: www.newscientist.com

6. Bollen and his colleagues wanted to find out _____.

- a) if the use of Twitter is on the rise
- b) how many different types of Twitter users there were
- c) why people of similar age, nationality and race are linked together
- d) whether degrees of happiness determine the choice of friends on Twitter

7. According to Bollen, which of the following could be an explanation for the behaviour of Twitter users?

- a) The tweets written may often be intended to heighten people's moods.
- b) People could be influenced by what they read on Twitter and reply accordingly.
- c) Some Twitter users tend to infect other users' PCs by spreading viruses.
- d) People may not express honest feelings in their tweets.

8. Which would be a suitable title for the text?

- a) Tweets Without Boundaries
- b) The Rising Popularity of Tweeter
- c) Your Tweets: Who do They Attract?
- d) Tweeter: A Threat to Social Relations?

(1) Biella, in northwest Italy, is the centre of a cluster of wool manufacturers and the home of Ermenegildo Zegna, a luxury clothing brand. A group of the town's businessmen have, however, come up with a scheme far removed from the catwalks and seasonal collections. They plan to use wool to soak up oil spills because it is good at repelling water and absorbing oil. The idea came to them last year, after the Deepwater Horizon disaster in the Gulf of Mexico, and it would, they think, have worked better than the chemical dispersants and other methods used there to clean up the oil spill.

(2) Earlier this year, Tecnomeccanica Biellese, an engineering firm that makes machinery for the wool industry, carried out experiments using coarse wool—the cheapest sort—to see how good it was at gathering oil. It turned out to be very good. In fact, it was able to absorb ten times its own weight in heavy fuel oil, a refinery product similar to crude oil. Moreover, the oil could be squeezed out and the wool reused. Indeed, even after a dozen immersions in oil, for between 15 and 20 seconds each time, the wool's absorptive capacity did not decline.

(3) Moving out of the laboratory and onto the water with a working oil-collection system is the next step. In March the businessmen, who have called their project the Wool Recycle Eco System, obtained patents for a containerised kit that can be set up on boats to deal with minor spills, and for a bigger, ship-based system to tackle serious ones. Mario Ploner, the managing director of Tecnomeccanica Biellese, says cleaning up the Deepwater Horizon spill of almost 5 million barrels would have required around 7,000 tonnes of wool. At the current market price of less than \$1 a kilo, that does not add up to a huge sum for a business as big as the oil industry. It would, on the other hand, be a nice little earner for sheep shearers.

Adapted from: www.twextra.com/aqgnrp

9. The businessmen from Biella have _____.

- a) proposed a scheme that can help fight environmental pollution
- b) devised a new method of recycling the waste wool from manufacturing plants
- c) claimed that wool could be employed as a chemical dispersant in the oceans
- d) invested in a new type of machine to revitalise the wool industry

10. It can be understood from the text that coarse wool _____.

- a) absorbs crude oil
- b) can soak up oil very quickly
- c) can only be used up to 20 times
- d) is able to absorb increasing amounts of oil after each use

11. Which is not true about the project Wool Recycle Eco System?

- a) It is cost-effective.
- b) It will benefit both the wool and the oil industries.
- c) It proved effective in the Deepwater Horizon disaster.
- d) It can be put to use in two different forms according to need.

(1) Washington state runs the largest ferry system in America. Each year 24 million commuters, tourists and others board the picturesque green-on-white boats and travel between islands in the state's north-west reaches, or across the miles-wide fjord called Puget Sound that faces the city of Seattle. But the ferry system has had some problems. Last year, four 80-year-old boats were removed from service because their hulls were rusting. That left small towns such as Port Townsend on the Olympic Peninsula virtually without tourists and shoppers. A small ferry borrowed from a nearby county has entered into service, but it is too flimsy to withstand the rough seas of winter and must often stay in port.

(2) But that was not the last of the bad news: one more ferry has been taken out of service on a route into Seattle, again because of rust; four more have been found to be in need of repair; two inadvertently crashed into the port; and one was hit by a large wave that submerged the front part, knocked out a watertight door and flooded the passenger area. Some of this is bad luck, but it is difficult to ignore the fragility of the system and the lack of care. At present there are no backup boats, which means the fleet **is being pushed to its limit**. Ferry officials hope to replace 18 of the current 21 boats over the next thirty years, but that is a long time and will require billions of dollars.

Adapted from:
www.ferriesbc.proboards.com

12. According to paragraph 1, which of the following is false?

- a) The borrowed boat has proved to be quite useful.
- b) Some boats in the ferry system need repairing.
- c) The sea conditions in winter are harsh.
- d) Last year almost no tourists could visit the small towns on the coast.

13. Which one is not a reason contributing to the collapse of the ferry system?

- a. Accidents
- b. Misfortune
- c. Poor maintenance
- d. Bad routes

14. In paragraph 2, "is being pushed to its limit" probably means _____.

- a) is being banned from service
- b) is the largest in the region
- c) is being forced to change its course
- d) is being worked beyond its capacity

(1) Many studies have found that well-educated people from developing countries are particularly likely to emigrate. This “brain drain” has long bothered policymakers in poor countries. They fear that it hurts their economies, as it deprives them of much-needed skilled workers who could teach at their universities, work in their hospitals and come up with clever new products for their factories to make.

(2) **Many now take issue with this view.** Several economists are of the opinion that the brain-drain hypothesis fails to account for the effects of remittances (money which is sent to someone), the beneficial effects of returning migrants, and the possibility that being able to migrate to developed countries stimulates people to get more education. They argue that once these factors are taken into account, the emigration of large numbers of highly skilled people could turn out to have a net benefit on the countries they leave. Recent studies of migration from countries as far apart as Ghana, Fiji, India and Romania have found support for this “brain gain” idea.

(3) It is true that many skilled migrants have been educated and trained partly at the expense of their governments. Some argue that these poor countries should therefore rethink how much they spend on higher education. But a new study of remittances sent home by Ghanaian migrants suggests that on average they transfer enough over their working lives to repay the amount spent on educating them several times over. The study finds that once remittances are taken into account, the cost of education would have to be 5 times higher than it is at present to be a financial burden on Ghana.

(4) There are more subtle ways in which the departure of some skilled people may aid poorer countries. Some emigrants would have been jobless had they stayed. Studies have found that unemployment rates among young people with college degrees in countries like Morocco and Tunisia are several multiples of **those** among the poorly educated, perhaps because graduates are more demanding. Migration may lead to a more productive pairing of people’s skills and jobs. Some of the benefits of this improved match then flow back to the migrant’s home country, most directly via remittances.

(5) The possibility of emigration may even have beneficial effects on those who choose to stay, by giving people in poor countries an incentive to invest in education. A study of Cape Verdeans finds that even a small increase in young people’s expectation of their emigrating increases the probability of their completing secondary school considerably. Another study looks at Fiji. A series of military coups beginning in 1987 was seen by Fijians of Indian origin as permanently harming their prospects in the country owing to the limits imposed on their share of government jobs and political power. This set off a wave of emigration. As a consequence, young Indians in Fiji became more likely to go to university even as the outlook at home dimmed, in part because Australia, Canada and New Zealand, three of the top destinations for Fijians, put more emphasis on attracting skilled migrants. Since some of those who got more education ended up staying, the skill levels of the resident Fijian population soared.

(6) Letting educated people go where they want looks like the clever option.

Adapted from: www.elearninghoje.blogspot.com

15. The writer does not mention that brain drain is feared to bring about a shortage in the number of _____ in poor countries.
- a) academicians
 - b) medical staff
 - c) top executives
 - d) inventors
16. In paragraph 2, "Many now take issue with this view" means that many people _____ that brain drain harms the economies of poor countries.
- a) support the idea
 - b) disagree with the idea
 - c) raise the issue
 - d) are aware
17. The writer argues that the remittances sent home by highly skilled migrants _____.
- a) increase the cost of education in their home countries in the long term
 - b) force their governments to invest less in education
 - c) compensate for the expenses their own governments have paid for their education
 - d) provide strong support for the brain-drain hypothesis
18. In paragraph 4, "those" refers to _____.
- a) studies
 - b) unemployment rates
 - c) countries
 - d) young people
19. Why does the writer mention the studies of Cape Verdeans and Fijians in Paragraph 5?
- a) To explain why some of the highly educated do not choose to emigrate
 - b) To prove that military coups can set off a wave of emigration among the highly skilled
 - c) To show how political power affects young people's desire to get education abroad
 - d) To illustrate that the prospect of emigration provides an incentive to get higher education
20. Which adjective best describes the writer's attitude toward the emigration of skilled workers from poor countries?
- a) Critical
 - b) Supportive
 - c) Biased
 - d) Impartial

THE END OF THE EXAM

EK-2: urGENOVA Temelli Çalışan G-String Betikleri ve Analiz Çıktıları

Dengelenmiş *bxm* Deseni Betiği ve Analiz Çıktıları

CONTROL CARDS FOR RUN 1
Control Cards File Name: ~Temp.txt
madde duzeyi

GSTUDY madde duzeyi
COMMENT
COMMENT Processing date: 12.11.2013 17:58:19
COMMENT G_String Version 6.3.3
COMMENT
COMMENT madde duzeyi
COMMENT
COMMENT_PATH: C:\Users\gulsen\genellenebilirlik\item_level\veri.txt
COMMENT
COMMENT% ogrenci (o)
COMMENT% madde (m)
OPTIONS NREC 5 "*.lis" TIME NOBANNER
EFFECT * o 1500
EFFECT m 20
FORMAT 30 0
PROCESS "~Temp.dat"

ANOVA TABLE FOR RUN 1
madde duzeyi

Effect	df	T	SS	MS	VC
o	1499	1305.54992	1305.54992	0.87095	0.03381
m	19	631.18063	631.18063	33.22003	0.02202
om	28481	7481.24955	5544.51900	0.19467	0.19467
Mean		0.00000			
Total	29999		7481.24955		
Grand Mean:		0.00000			

Facets with adjusted sample sizes according to Brennan

'o' Differentiation 1500,00
'm' Random 20,00

Pattern	Var. Comp.	Levels	Signature	Rule
o	0,034	1,00	d	tau only
m	0,001	20,00	r	Delta only
o m	0,010	20,00	dr	Delta and delta

RESULTS:

s2(T) = 0,034
s2(D) = 0,011
s2(d) = 0,010
Er2 = 0,776
Phi = 0,757

Dengelenmiş *bxt* Deseni Betiği ve Analiz Çıktıları

CONTROL CARDS FOR RUN 1
Control Cards File Name: ~Temp.txt
madde takimi duzeyi

GSTUDY madde takimi duzeyi
COMMENT
COMMENT Processing date: 12.11.2013 18:01:29
COMMENT G_String Version 6.3.3
COMMENT
COMMENT madde takimi duzeyi
COMMENT
COMMENT_PATH: C:\Users\gulsen\genellenebilirlik\testlet_level\testlet_veri.txt
COMMENT
COMMENT% ogrenci (o)
COMMENT% madde takimi (t)
OPTIONS NREC 5 "*.lis" TIME NOBANNER
EFFECT * o 1500
EFFECT t 6
FORMAT 30 0
PROCESS "~Temp.dat"

ANOVA TABLE FOR RUN 1
madde takimi duzeyi

Effect	df	T	SS	MS	VC
o	1499	423.51838	423.51837	0.28253	0.03586
t	5	147.81910	147.81909	29.56382	0.01966
ot	7495	1076.13706	504.79960	0.06735	0.06735
Mean		0.00001			
Total	8999		1076.13706		
Grand Mean:		0.00003			

Facets with adjusted sample sizes according to Brennan

'o' Differentiation 1500,00
't' Random 6,00

Pattern	Var. Comp.	Levels	Signature	Rule
o	0,036	1,00	d	tau only
t	0,003	6,00	r	Delta only
o t	0,011	6,00	dr	Delta and delta

RESULTS:

s2(T) = 0,036
s2(D) = 0,015
s2(d) = 0,011
Er2 = 0,762
Phi = 0,712

Dengelenmemiş $bx(m:t)$ Deseni Betiği ve Analiz Çıktıları

CONTROL CARDS FOR RUN 1
Control Cards File Name: ~Temp.txt
madde takimi etkisi dahil

GSTUDY madde takimi etkisi dahil
COMMENT
COMMENT Processing date: 12.11.2013 18:10:50
COMMENT G_String Version 6.3.3
COMMENT
COMMENT
COMMENT_PATH: C:\Users\gulsen\genellenebilirlik\item_testlet_nested\veri.txt
COMMENT
COMMENT% ogrenci (o)
COMMENT% madde takimi (t)
COMMENT% madde (m)
OPTIONS NREC 5 "*.lis" TIME NOBANNER
EFFECT * o 1500
EFFECT t 6
EFFECT m:t 2 3 3 3 3 6
FORMAT 30 0
PROCESS "~Temp.dat"

ANOVA TABLE FOR RUN 1
madde takimi etkisi dahil

Effect	df	T	SS	MS	VC
o	1499	1305.54992	1305.54992	0.87095	0.03314
t	5	403.12864	403.12864	80.62573	0.01323
m:t	14	631.18063	228.05199	16.28943	0.01073
ot	7495	3249.58314	1540.90457	0.20559	0.00457
om:t	20986	7481.24955	4003.61443	0.19078	0.19078
Mean		0.00000			
Total	29999		7481.24955		
Grand Mean:		0.00000			

Facets with adjusted sample sizes according to Brennan

'o'	Differentiation	1500,00
't'	Random	5,26
'm'	Random	3,33

Pattern	Var. Comp.	Levels	Signature	Rule
o	0,033	1,00	d	tau only
t	0,003	5,26	r	Delta only
m:t	0,001	20,00	r	Delta only
o t	0,001	5,26	dr	Delta and delta
o m:t	0,010	20,00	dr	Delta and delta

RESULTS:

s²(T) = 0,033
s²(D) = 0,013
s²(d) = 0,010
Er² = 0,761
Phi = 0,711

EK-3: WinBUGS Betikleri ve Analiz Çıktıları

MTK Tabanlı DMF Kestiriminde Kullanılan WinBUGS Betiği

```
model
{
mu.Beta1<-mean(Beta1[])
mu.z<-mean(z[])
sd.z<-sd(z[])
for (i in 1:n){
a[i]~dnorm(mua,siga) l(0,)
b[i]~dnorm(mub,sigb)
Beta1[i]~dnorm(muBeta1,sigBeta1)
adj.Beta1[i]<-(Beta1[i]-mu.Beta1)
}
mua~dnorm(0,.0001)
mub~dnorm(0,.0001)
Beta~dnorm(muBeta,sigBeta)
muBeta~dnorm(0,.0001)
muBeta1~dnorm(0,.0001)
siga~dchisqr(.5)
sigb~dchisqr(.5)
sigBeta~dchisqr(.5)
sigBeta1~dchisqr(.5)
for (j in 1:N){
for (i in 1:n){
p[j,i]<- 1/(1+exp(-(a[i]*(Beta*Group[j]+z[j])-b[i]-Beta1[i]*Group[j])))
U[j,i]~dbern(p[j,i])
}
z[j] ~ dnorm(0,1)
}
adj.Beta<-(Beta-mu.Beta1)
}
```

Madde Takımları İçin İki Faktörlü ÇBMTK'ya Dayalı Madde takımlarında DMF Kestiriminde Kullanılan WinBUGS Betiği

```
model
{
  mu.Beta<-mean(Beta[])
  BetaT~dnorm(muBetaT,sigBetaT)
  mu.z<-mean(z[])
  sd.z<-sd(z[])
  for (i in 1:n){
    a[i]~dnorm(mua,siga) l(0,)
    b[i]~dnorm(mub,sigb)
    Beta[i]~dnorm(muBeta,sigBeta)
    adj.Beta[i]<-(Beta[i]-mu.Beta)
  }
  mua~dnorm(0,.0001)
  mub~dnorm(0,.0001)
  muBetaT~dnorm(0,.0001)
  muBeta~dnorm(0,.0001)
  siga~dchisqr(.5)
  sigb~dchisqr(.5)
  sigBetaT~dchisqr(.5)
  sigBeta~dchisqr(.5)
  for (j in 1:N){
    for (i in 1:n){
      p[j,i]<- 1/(1+exp((-a[i]*(BetaT*Group[j]+z[j]-b[i]+gamma[j,i]-
Beta[i]*Group[j])))
      U[j,i]~dbern(p[j,i])
      gamma[j,i]<-gamtes[j,test[i]]
    }
    z[j]~dnorm(0,1)
    for (i in 1:NT){
      gamtes[j,i]~dnorm(0,sigg[i])
    }
  }
  adj.BetaT<-(BetaT-mu.Beta)
  for (i in 1:NT){
    sigg[i]~dchisqr(.5)
    varg[i]<-1/sigg[i]
  }
}
```

MTK Tabanlı DMF Kestirimi WinBUGS Sonucu: *Cinsiyet*

Node
statistics

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
Beta	182,3449	0.2981587	0.03934787	1.817.977	1.823.797	1.828.697	3001	6000
Beta1[1]	182,3472	0.3120882	0.03981819	1.817.781	1.823.636	1.829.099	3001	6000
Beta1[2]	182,3617	0.3118939	0.03995878	1.817.957	1.823.861	1.829.444	3001	6000
Beta1[3]	182,3684	0.3133021	0.0401405	1.817.868	1.823.861	1.829.686	3001	6000
Beta1[4]	182,7549	0.3074784	0.03943412	1.821.928	1.827.731	1.833.232	3001	6000
Beta1[5]	182,3775	0.3131487	0.03933291	1.818.028	1.823.956	1.829.618	3001	6000
Beta1[6]	181,6988	0.319938	0.03990171	181.093	1.817.145	1.822.874	3001	6000
Beta1[7]	181,8747	0.3162868	0.03981969	1.812.848	1.818.966	1.824.673	3001	6000
Beta1[8]	181,9952	0.3162005	0.03934324	1.814.176	1.820.131	1.825.834	3001	6000
Beta1[9]	182,2814	0.3226211	0.03940272	1.816.733	182.297	1.828.977	3001	6000
Beta1[10]	183,0545	0.3359027	0.03970997	1.824.157	1.830.575	1.836.878	3001	6000
Beta1[11]	181,8494	0.3246397	0.03972459	1.810.445	1.816.677	1.822.508	3001	6000
Beta1[12]	182,3000	0.3226301	0.04007652	1.817.026	1.823.203	1.829.259	3001	6000
Beta1[13]	182,0990	0.3195023	0.03951491	1.814.799	1.821.272	1.826.738	3001	6000
Beta1[14]	181,8138	0.3279195	0.03979869	182.213	1.828.197	1.834.495	3001	6000
Beta1[15]	182,1304	0.3247862	0.03947734	1.815.221	1.821.451	1.827.358	3001	6000
Beta1[16]	182,0526	0.3052055	0.03930624	1.814.899	1.820.771	1.826.118	3001	6000
Beta1[17]	182,4890	0.3072267	0.0397348	1.819.249	1.825.153	1.830.536	3001	6000
Beta1[18]	182,5149	0.3046054	0.039044	1.819.621	1.825.417	1.830.817	3001	6000
Beta1[19]	182,3063	0.3109718	0.03928216	181.729	1.823.298	1.828.914	3001	6000
Beta1[20]	182,1056	0.3130703	0.03942269	1.815.353	182.124	1.826.796	3001	6000

Madde Takımları İçin İki Faktörlü ÇBMTK'ya Dayalı Madde Takımlarında DMF Kestirimi WinBUGS Sonucu: *Cinsiyet*

Node
statistics

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
Beta[1]	37,23034	0.2239092	0.03110888	3.680.502	3.720.976	3.763.783	3001	6000
Beta[2]	37,24554	0.2188686	0.03072571	3.681.575	3.723.104	3.764.407	3001	6000
Beta[3]	37,2513	0.2231578	0.03139459	3.681.077	3.723.334	3.766.068	3001	6000
Beta[4]	37,60786	0.2251555	0.03186179	3.717.654	37.58	3.802.177	3001	6000
Beta[5]	37,25406	0.2236982	0.03137878	368.305	3.724.019	376.702	3001	6000
Beta[6]	36,45103	0.269667	0.03507437	3.593.423	3.642.615	3.696.132	3001	6000
Beta[7]	36,6589	0.2545269	0.0326813	361.822	3.663.466	3.714.616	3001	6000
Beta[8]	36,78961	0.2649104	0.03289563	3.627.887	367.776	3.731.237	3001	6000
Beta[9]	37,15882	0.2632711	0.03132221	366.455	3.714.392	3.766.962	3001	6000
Beta[10]	38,06478	0.286914	0.03291767	3.751.038	3.806.103	3.861.281	3001	6000
Beta[11]	36,47502	0.2614954	0.03273916	359.812	3.647.021	3.697.955	3001	6000
Beta[12]	37,16656	0.2486374	0.0324142	3.669.206	3.716.209	3.763.276	3001	6000
Beta[13]	36,94363	0.2559034	0.033276	3.645.098	3.693.147	3.744.227	3001	6000
Beta[14]	36,38537	0.3712906	0.03146727	3.735.583	3.810.297	3.881.732	3001	6000
Beta[15]	36,88115	0.3787785	0.0325702	3.612.626	368.821	3.760.696	3001	6000
Beta[16]	36,79843	0.2425428	0.03216382	363.241	3.680.227	3.724.747	3001	6000
Beta[17]	37,46755	0.2415424	0.03245729	3.701.415	3.745.163	3.792.447	3001	6000
Beta[18]	37,49592	0.2408318	0.03140773	3.704.755	3.748.616	3.794.667	3001	6000
Beta[19]	37,15542	0.2451689	0.0313468	3.668.538	3.714.791	3.759.296	3001	6000
Beta[20]	36,92932	0.2561452	0.03224549	3.643.261	3.693.803	3.740.019	3001	6000
BetaT	37,24247	0.216434	0.03141817	368.474	3.723.843	3.762.584	3001	6000

MTK Tabanlı DMF Kestirimi WinBUGS Sonucu: *Bölüm (Sayısal-Eşit Ağırlık)*

Node
statistics

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
Beta	-107,9898	0.570504	0.057883	-1.093.811	-108.048	-106.558	15006	7995
Beta[1]	-107,7869	0.538594	0.0577046	-1.092.334	-1.078.328	-1.062.999	15006	7995
Beta[2]	-107,4958	0.552847	0.0575428	-108.937	-1.075.297	-1.060.326	15006	7995
Beta[3]	-107,9344	0.560727	0.0584455	-1.094.072	-107.965	-1.064.397	15006	7995
Beta[4]	-107,6118	0.551332	0.0571254	-1.090.536	-107.641	-1.061.088	15006	7995
Beta[5]	-107,6957	0.551432	0.0576534	-1.091.666	-1.077.299	-1.061.898	15006	7995
Beta[6]	-108,134	0.563814	0.0577706	-1.093.064	-1.078.738	-1.063.373	15006	7995
Beta[7]	-108,152	0.567544	0.0580762	-1.092.767	-1.078.698	-1.063.529	15006	7995
Beta[8]	-107,9813	0.5617	0.057754	-1.094.487	-1.080.135	-1.064.812	15006	7995
Beta[9]	-108,2517	0.569537	0.0579262	-1.097.413	-1.082.742	-1.067.484	15006	7995
Beta[10]	-107,8513	0.57226	0.0578725	-1.093.375	-1.078.752	-1.063.183	15006	7995
Beta[11]	-108,014	0.579162	0.0580279	-1.094.349	-1.079.881	-1.064.897	15006	7995
Beta[12]	-107,8956	0.577226	0.0583013	-1.093.678	-1.079.215	-106.422	15006	7995
Beta[13]	-107,8353	0.565518	0.0579318	-1.093.111	-1.078.613	-1.063.255	15006	7995
Beta[14]	-107,5824	0.574585	0.0581056	-1.090.941	-1.076.059	-1.060.255	15006	7995
Beta[15]	-107,6159	0.56526	0.0576664	-1.091.637	-1.076.326	-1.060.217	15006	7995
Beta[16]	-108,1512	0.562072	0.0572599	-1.096.076	-1.081.753	-1.066.501	15006	7995
Beta[17]	-107,956	0.560308	0.058058	-1.094.343	-1.079.793	-1.064.686	15006	7995
Beta[18]	-107,8516	0.553678	0.0575278	-109.307	-1.078.736	-1.063.443	15006	7995
Beta[19]	-107,8873	0.552553	0.0575933	-1.093.763	-1.079.172	-1.063.969	15006	7995
Beta[20]	-107,828	0.561736	0.058116	-1.092.988	-1.078.483	-1.063.192	15006	7995

Madde Takımları İçin İki Faktörlü ÇBMTK'ya Dayalı Madde Takımlarında DMF Kestirimi WinBUGS Sonucu: Bölüm (Sayısal-Eşit Ağırlık)

Node statistics

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
Beta1[1]	182,3472	0.3120882	0.03981819	1.817.781	1.823.636	1.829.099	3001	6000
Beta1[2]	182,7809	0.3118939	0.03995878	1.817.957	1.823.861	1.829.444	3001	6000
Beta1[3]	182,3684	0.3133021	0.0401405	1.817.868	1.823.861	1.829.686	3001	6000
Beta1[4]	182,7549	0.3074784	0.03943412	1.821.928	1.827.731	1.833.232	3001	6000
Beta1[5]	182,3775	0.3131487	0.03933291	1.818.028	1.823.956	1.829.618	3001	6000
Beta1[6]	181,6988	0.319938	0.03990171	181.093	1.817.145	1.822.874	3001	6000
Beta1[7]	181,8747	0.3162868	0.03981969	1.812.848	1.818.966	1.824.673	3001	6000
Beta1[8]	181,9952	0.3162005	0.03934324	1.814.176	1.820.131	1.825.834	3001	6000
Beta1[9]	182,2814	0.3226211	0.03940272	1.816.733	182.297	1.828.977	3001	6000
Beta1[10]	183,0545	0.3359027	0.03970997	1.824.157	1.830.575	1.836.878	3001	6000
Beta1[11]	181,6494	0.3246397	0.03972459	1.810.445	1.816.677	1.822.508	3001	6000
Beta1[12]	182,3	0.3226301	0.04007652	1.817.026	1.823.203	1.829.259	3001	6000
Beta1[13]	182,099	0.3195023	0.03951491	1.814.799	1.821.272	1.826.738	3001	6000
Beta1[14]	182,8138	0.3279195	0.03979869	182.213	1.828.197	1.834.495	3001	6000
Beta1[15]	182,1304	0.3247862	0.03947734	1.815.221	1.821.451	1.827.358	3001	6000
Beta1[16]	182,0526	0.3052055	0.03930624	1.814.899	1.820.771	1.826.118	3001	6000
Beta1[17]	182,489	0.3072267	0.0397348	1.819.249	1.825.153	1.830.536	3001	6000
Beta1[18]	182,5149	0.3046054	0.039044	1.819.621	1.825.417	1.830.817	3001	6000
Beta1[19]	182,3063	0.3109718	0.03928216	181.729	1.823.298	1.828.914	3001	6000
Beta1[20]	182,1056	0.3130703	0.03942269	1.815.353	182.124	1.826.796	3001	6000
BetaT	182,3449	0.2981587	0.03934787	1.817.977	1.823.797	1.828.697	3001	6000

EK-4: Örnek Uzman Kanısı Formu

Sayın Katılımcı,

Aşağıda, bir devlet üniversitesinin Yabancı Diller Yüksek Okulu tarafından yapılan İngilizce Yeterlik Sınavı'nda yer alan bazı sorular ile ilgili görüşünüze başvurulacaktır. Doktora tez çalışması için kullanılan bu sorulara yönelik değişik alt gruplardan (cinsiyet - bölüm) oluşan öğrencilerin bu sorulara yönelik performansları **DMF** analizleri ile incelenmektedir.

DMF: Eşit yetenek düzeyinde ancak farklı gruplarda yer alan bireylerin bir maddedeki (sorudaki) performanslarının farklı olması değişen madde fonksiyonu (DMF) ile açıklanmaktadır. Bu durum **madde yanlılığının** bir göstergesidir.

Testle ölçülen özellik açısından aynı yetenek düzeyinde olan ancak cinsiyet, sosyo-ekonomik düzey gibi değişkenler açısından birbirinden farklı alt gruplarda yer alan bireylerin, bir maddeyi doğru cevaplandırma olasılıklarının farklılaşması olarak da tanımlanabilir. Örneğin bir sayısal yetenek probleminde problemin konusu basketbol, futbol, araba yarışları veya avcılık gibi daha çok erkeklerin ilgisini çeken konulardan seçildiğinde, maddeleri (aynı yetenek düzeyine sahip olsalar bile) erkeklerin doğru cevaplama olasılığı artmaktadır. Benzer şekilde, kız öğrencilerin daha aşina olduğu konuların yer aldığı maddeler ise kız öğrencilerin lehine işlemektedir.

Maddelerin hangi özelliklerinin DMF'ye yol açtığı belirlenmesi için uzman görüşlerine ihtiyaç duyulmaktadır. Sizden istenen aşağıda verilen her bir okuma parçasına ilişkin DMF gösteren sorulardaki yanlılık ihtimalinin neden kaynaklanmış olabileceğine dair uzman görüşünüzü belirtmenizdir.

Vereceğiniz cevaplar sadece "uzman görüşü" kapsamında, belirtilen çalışmada kullanılacaktır.

Desteyiniz için teşekkür ederim.

Gülşen TAŞDELEN TEKER

1. OKUMA PARÇASI

(1) Last week, a pair of Russian space modules arrived on the Isle of Man, in the British Isles, home to the private space company Excalibur. The firm will work on updating the modules, and plans eventually to use them to provide extra room and supplies for the tourists and researchers it hopes to take into space. "They are not the nice shiny stuff, but they have been proven to be a reliable means of space travel," says former astronaut Leroy Chiao, who now works for Excalibur.

(2) The modules, which have never been launched, were built in the 1970s as part of Almaz, a military programme that sent cosmonauts into orbit to take photographs of the Earth. Excalibur has also bought four Almaz spacecraft, which could be launched sooner than the modules, to start ferrying NASA astronauts into space.

<p>2. Excalibur _____.</p> <p>a) employs a former astronaut</p> <p>b) has been carrying NASA astronauts into space</p> <p>c) was actively involved in Almaz</p> <p>d) has recently launched four spacecraft into orbit</p>	<p>Yandaki maddenin sayısal-eşit ağırlık karşılaştırmasında eşit ağırlık bölümü öğrencilerin lehine işlediği yapılan analiz sonucu ortaya konmuştur. Bu maddenin yanlı çıkmasına katılma durumunuzu belirtiniz.</p> <p>Kesinlikle Katılıyorum ()</p> <p>Katılıyorum (x)</p> <p>Katılmıyorum ()</p> <p>Cevabınızın gerekçesini belirtiniz.</p> <p><i>Paragraf sorularında sözel yeteneğe sahip olan bireyler kelimeleri daha etkili kullanmakta ve kelimeleri, dilleri daha kolay öğrenmektedirler. Eşit ağırlık bölümü öğrencilerinin sayısal bölüm öğrencilerine göre bu soruda daha başarılı olması beklenebilir.</i></p>
<p>Doğru cevap: a</p>	<p>Yandaki maddenin sayısal-sözel karşılaştırmasında sözel bölüm öğrencilerin lehine işlediği yapılan analiz sonucu ortaya konmuştur. Bu maddenin yanlı çıkmasına katılma durumunuzu belirtiniz.</p> <p>Kesinlikle Katılıyorum ()</p> <p>Katılıyorum (X)</p> <p>Katılmıyorum ()</p> <p>Cevabınızın gerekçesini belirtiniz.</p> <p><i>Yukarıda yazdığım gerekçe bu iki alanın karşılaştırılmasında da geçerli.</i></p>

2. OKUMA PARÇASI

(1) Social scientists have long known that social networks tend to have an "assortative" character—people are more likely to have friends who are similar in age, nationality or race, and less likely to be linked to people who are different. Psychologist Johan Bollen of the University of Indiana and colleagues wondered if people's moods is also a factor which draws them together on an online network like Twitter.

(2) To find out, they tracked 102,000 Twitter users over six months, analysing their 129 million, 140-letter-or-less tweets, using standard techniques from psychology. Specifically, they measured the emotional content of the tweets as reflected by the presence of positive or negative words. From this they could assess the "subjective well-being", or happiness, of the users through their tweets. The researchers indeed found that happier people—those using positive statements—tended to be tweeting and receiving tweets from people who were also happier. Similarly, less happy people tended to group together. According to Bollen, Twitter users are preferentially linked to those with whom they share a similar level of general happiness.

(3) Bollen admits they don't yet know why this is true. Happy or unhappy people may simply seek one another out, drawn by tweets expressing emotions similar to their own. Or, Bollen suggests, it could be that the emotions expressed, even in short tweets, have an infectious quality, lifting peoples' spirits or filling them with gloom, depending on what they read.

<p>6. Bollen and his colleagues wanted to find out _____.</p> <p>a) if the use of Twitter is on the rise</p> <p>b) how many different types of Twitter users there were</p> <p>c) why people of similar age, nationality and race are linked together</p> <p>d) whether degrees of happiness determine the choice of friends on Twitter</p> <p>Doğru cevap: d</p>	<p>Yandaki maddenin erkekler lehine işlediği yapılan analiz sonucu ortaya konmuştur. Bu maddenin yanlı çıkmasına katılma durumunuzu belirtiniz.</p> <p>Kesinlikle Katılıyorum() Katılıyorum () Katılmıyorum (X)</p> <p>Cevabınızın gerekçesini belirtiniz.</p> <p><i>Bu maddenin kadınlar lehine yanlı çıkmasını bekliyordum. Kadınların twitter kullanımının daha yüksek olması bu soruda kadınlar lehine bir yanlılık yaratabilir diye düşünüyorum.</i></p> <p>Yandaki maddenin sayısal-eşit ağırlık karşılaştırmasında sayısal bölüm öğrencilerin lehine işlediği yapılan analiz sonucu ortaya konmuştur. Bu maddenin yanlı çıkmasına katılma durumunuzu belirtiniz.</p> <p>Kesinlikle Katılıyorum() Katılıyorum () Katılmıyorum (X)</p> <p>Cevabınızın gerekçesini belirtiniz.</p> <p><i>Bu soruda yapılan bölüm karşılaştırmasına göre yanlılık çıkmamasını bekliyorum. Bir yanlılık olacaksa da psikoloji bölümü ile ilgili bir konu olduğu için eşit ağırlık öğrencilerinin lehine çıkmasını bekliyordum.</i></p>
<p>7. According to Bollen, which of the following could be an explanation for the behaviour of Twitter users?</p> <p>a) The tweets written may often be intended to heighten people's moods.</p> <p>b) People could be influenced by what they read on Twitter and reply accordingly.</p> <p>c) Some Twitter users tend to infect other users' PCs by spreading viruses.</p> <p>d) People may not express honest feelings in their tweets.</p> <p>Doğru cevap: b</p>	<p>Yandaki maddenin erkekler lehine işlediği yapılan analiz sonucu ortaya konmuştur. Bu maddenin yanlı çıkmasına katılma durumunuzu belirtiniz.</p> <p>Kesinlikle Katılıyorum() Katılıyorum () Katılmıyorum (x)</p> <p>Cevabınızın gerekçesini belirtiniz.</p> <p><i>Kadınların lehine yanlılık bekliyordum. Bunun nedeni 1- twitter kullanım oranının ve kullanım süresinin kadınlarda daha yüksek olması, 2- kadınların eleştirel düşünme becerilerinin erkeklere göre daha iyi olması. Bu soruda davranışın açıklanmasına ilişkin olduğundan dolayı eleştirel düşünme becerilerinin etkisi olacağını düşünüyorum.</i></p> <p>Yandaki maddenin sayısal-eşit ağırlık karşılaştırmasında sayısal bölüm öğrencilerin lehine işlediği yapılan analiz sonucu ortaya konmuştur. Bu maddenin yanlı çıkmasına katılma durumunuzu belirtiniz.</p> <p>Kesinlikle Katılıyorum () Katılıyorum (x) Katılmıyorum ()</p> <p>Cevabınızın gerekçesini belirtiniz.</p> <p><i>Eleştirel düşünme becerileri mantıksal ve matematiksel zeka alanının altında yer aldığından ötürü, sayısal öğrencilerin bunu yakalayabilmeleri onların lehine yanlı çıkmasının nedeni olabilir.</i></p>
<p>8. Which would be a suitable title for the text?</p> <p>a) Tweets Without Boundaries</p> <p>b) The Rising Popularity of Tweeter</p> <p>c) Your Tweets: Who do They Attract?</p> <p>d) Tweeter: A Threat to Social Relations?</p> <p>Doğru cevap: c</p>	<p>Yandaki maddenin erkekler lehine işlediği yapılan analiz sonucu ortaya konmuştur. Bu maddenin yanlı çıkmasına katılma durumunuzu belirtiniz.</p> <p>Kesinlikle Katılıyorum() Katılıyorum () Katılmıyorum (x)</p> <p>Cevabınızın gerekçesini belirtiniz.</p> <p><i>Cinsiyet açısından yanlılık beklemiyordum.</i></p>

3. OKUMA PARÇASI

(1) Biella, in northwest Italy, is the centre of a cluster of wool manufacturers and the home of Ermenegildo Zegna, a luxury clothing brand. A group of the town's businessmen have, however, come up with a scheme far removed from the catwalks and seasonal collections. They plan to use wool to soak up oil spills because it is good at repelling water and absorbing oil. The idea came to them last year, after the Deepwater Horizon disaster in the Gulf of Mexico, and it would, they think, have worked better than the chemical dispersants and other methods used there to clean up the oil spill.

(2) Earlier this year, Tecnomeccanica Biellese, an engineering firm that makes machinery for the wool industry, carried out experiments using coarse wool—the cheapest sort—to see how good it was at gathering oil. It turned out to be very good. In fact, it was able to absorb ten times its own weight in heavy fuel oil, a refinery product similar to crude oil. Moreover, the oil could be squeezed out and the wool reused. Indeed, even after a dozen immersions in oil, for between 15 and 20 seconds each time, the wool's absorptive capacity did not decline.

(3) Moving out of the laboratory and onto the water with a working oil-collection system is the next step. In March the businessmen, who have called their project the Wool Recycle Eco System, obtained patents for a containerised kit that can be set up on boats to deal with minor spills, and for a bigger, ship-based system to tackle serious ones. Mario Ploner, the managing director of Tecnomeccanica Biellese, says cleaning up the Deepwater Horizon spill of almost 5 million barrels would have required around 7,000 tonnes of wool. At the current market price of less than \$1 a kilo, that does not add up to a huge sum for a business as big as the oil industry. It would, on the other hand, be a nice little earner for sheep shearers.

<p>10. It can be understood from the text that coarse wool _____.</p> <p>a) absorbs crude oil b) can soak up oil very quickly c) can only be used up to 20 times d) is able to absorb increasing amounts of oil after each use</p> <p>Doğru cevap: b</p>	<p>Yandaki maddenin kadınlar lehine işlediği yapılan analiz sonucu ortaya konmuştur. Bu maddenin yanlı çıkmasına katılma durumunuzu belirtiniz.</p> <p>Kesinlikle Katılıyorum (X) Katılıyorum () Katılmıyorum ()</p> <p>Cevabınızın gerekçesini belirtiniz. <i>Kadınların kıyafet dokuları, kumaş vb. Konularda erkeklere göre daha fazla aşına olduğunu düşünüyorum.</i></p> <p>Yandaki maddenin sayısal-eşit ağırlık karşılaştırmasında eşit ağırlık bölümü öğrencilerin lehine işlediği yapılan analiz sonucu ortaya konmuştur. Bu maddenin yanlı çıkmasına katılma durumunuzu belirtiniz.</p> <p>Kesinlikle Katılıyorum () Katılıyorum (x) Katılmıyorum ()</p> <p>Cevabınızın gerekçesini belirtiniz. <i>Paragraf sorularında, alana özel bir bilgi yoksa (ki burada öyle bir durum fark etmedim) sözel yeteneğe sahip olan bireyler kelimeleri daha etkili kullanmakta ve kelimeleri, dilleri daha kolay öğrenmektedirler. Eşit ağırlık bölümü öğrencilerinin sayısal bölüm öğrencilerine göre bu soruda daha başarılı olması beklenebilir.</i></p>
<p>11.Which is <u>not</u> true about the project Wool Recycle Eco System?</p> <p>a) It is cost-effective. b) It will benefit both the wool and the oil industries. c) It proved effective in the Deepwater Horizon disaster. d) It can be put to use in two different forms according to need.</p> <p>Doğru cevap: c</p>	<p>Yandaki maddenin erkekler lehine işlediği yapılan analiz sonucu ortaya konmuştur. Bu maddenin yanlı çıkmasına katılma durumunuzu belirtiniz.</p> <p>Kesinlikle Katılıyorum () Katılıyorum (X) Katılmıyorum ()</p> <p>Cevabınızın gerekçesini belirtiniz. <i>erkekler lehine yanlı çıkması bir sistem söz konusu olunca mantıklı geliyor. Makine, endüstri ve mühendislik ile ilgili kavramlar çok fazla yer aldığından dolayı erkeklerin daha aşına olacağını düşünüyorum.</i></p> <p>Yandaki maddenin sayısal-eşit ağırlık karşılaştırmasında sayısal bölümü öğrencilerin lehine işlediği yapılan analiz sonucu ortaya konmuştur. Bu maddenin yanlı çıkmasına katılma durumunuzu belirtiniz.</p> <p>Kesinlikle Katılıyorum () Katılıyorum (x) Katılmıyorum ()</p> <p>Cevabınızın gerekçesini belirtiniz. <i>Sayısal öğrencilerinin yukarıda sıraladığım makine, kimyasallar, endüstri, mühendislik gibi kavramlara daha aşın olacağını düşünüyorum. Sayısal Bölümler ile ilişkili konular olduğundan dolayı parçayı daha rahat kavrayabildiklerini düşünüyorum</i></p>

4. OKUMA PARÇASI

(1) Washington state runs the largest ferry system in America. Each year 24 million commuters, tourists and others board the picturesque green-on-white boats and travel between islands in the state's north-west reaches, or across the miles-wide fjord called Puget Sound that faces the city of Seattle. But the ferry system has had some problems. Last year, four 80-year-old boats were removed from service because their hulls were rusting. That left small towns such as Port Townsend on the Olympic Peninsula virtually without tourists and shoppers. A small ferry borrowed from a nearby county has entered into service, but it is too flimsy to withstand the rough seas of winter and must often stay in port.

(2) But that was not the last of the bad news: one more ferry has been taken out of service on a route into Seattle, again because of rust; four more have been found to be in need of repair; two inadvertently crashed into the port; and one was hit by a large wave that submerged the front part, knocked out a watertight door and flooded the passenger area. Some of this is bad luck, but it is difficult to ignore the fragility of the system and the lack of care. At present there are no backup boats, which means the fleet **is being pushed to its limit**. Ferry officials hope to replace 18 of the current 21 boats over the next thirty years, but that is a long time and will require billions of dollars

<p>14. In paragraph 2, "is being pushed to its limit" probably means _____.</p> <p>a) is being banned from service b) is the largest in the region c) is being forced to change its course d) is being worked beyond its capacity</p> <p>Doğru cevap: d</p>	<p>Yandaki maddenin erkekler lehine işlediği yapılan analiz sonucu ortaya konmuştur. Bu maddenin yanlı çıkmasına katılma durumunuzu belirtiniz.</p> <p>Kesinlikle Katılıyorum () Katılıyorum (x) Katılmıyorum ()</p> <p>Cevabınızın gerekçesini belirtiniz.</p> <p><i>"Boats", aynı satırda "back up" kelimesi ve parçanın genelinin deniz ulaşımı ile ilgili olması erkeklerin lehine yanlı olmasının nedeni olabilir.</i></p> <p>Yandaki maddenin sayısal-eşit ağırlık karşılaştırmasında eşit ağırlık bölümü öğrencilerin lehine işlediği yapılan analiz sonucu ortaya konmuştur. Bu maddenin yanlı çıkmasına katılma durumunuzu belirtiniz.</p> <p>Kesinlikle Katılıyorum () Katılıyorum (x) Katılmıyorum ()</p> <p>Cevabınızın gerekçesini belirtiniz.</p> <p><i>Uluslararası ticaret, uluslararası işletme, işletme, ekonomi vb bölümlerde okuyan öğrencilerin ilgi alanları ile bağlantılı olduğundan ve bu bölümlere de eşit ağırlık puanı ile öğrenci alındığından, bu öğrencilerin aşına olması olasıdır.</i></p>
---	---

5. OKUMA PARÇASI

(1) Many studies have found that well-educated people from developing countries are particularly likely to emigrate. This “brain drain” has long bothered policymakers in poor countries. They fear that it hurts their economies, as it deprives them of much-needed skilled workers who could teach at their universities, work in their hospitals and come up with clever new products for their factories to make.

(2) **Many now take issue with this view.** Several economists are of the opinion that the brain-drain hypothesis fails to account for the effects of remittances (money which is sent to someone), the beneficial effects of returning migrants, and the possibility that being able to migrate to developed countries stimulates people to get more education. They argue that once these factors are taken into account, the emigration of large numbers of highly skilled people could turn out to have a net benefit on the countries they leave. Recent studies of migration from countries as far apart as Ghana, Fiji, India and Romania have found support for this “brain gain” idea.

(3) It is true that many skilled migrants have been educated and trained partly at the expense of their governments. Some argue that these poor countries should therefore rethink how much they spend on higher education. But a new study of remittances sent home by Ghanaian migrants suggests that on average they transfer enough over their working lives to repay the amount spent on educating them several times over. The study finds that once remittances are taken into account, the cost of education would have to be 5 times higher than it is at present to be a financial burden on Ghana.

(4) There are more subtle ways in which the departure of some skilled people may aid poorer countries. Some emigrants would have been jobless had they stayed. Studies have found that unemployment rates among young people with college degrees in countries like Morocco and Tunisia are several multiples of **those** among the poorly educated, perhaps because graduates are more demanding. Migration may lead to a more productive pairing of people’s skills and jobs. Some of the benefits of this improved match then flow back to the migrant’s home country, most directly via remittances.

(5) The possibility of emigration may even have beneficial effects on those who choose to stay, by giving people in poor countries an incentive to invest in education. A study of Cape Verdeans finds that even a small increase in young people’s expectation of their emigrating increases the probability of their completing secondary school considerably. Another study looks at Fiji. A series of military coups beginning in 1987 was seen by Fijians of Indian origin as permanently harming their prospects in the country owing to the limits imposed on their share of government jobs and political power. This set off a wave of emigration. As a consequence, young Indians in Fiji became more likely to go to university even as the outlook at home dimmed, in part because Australia, Canada and New Zealand, three of the top destinations for Fijians, put more emphasis on attracting skilled migrants. Since some of those who got more education ended up staying, the skill levels of the resident Fijian population soared.

(6) Letting educated people go where they want looks like the clever option.

16. In paragraph 2, "Many now take issue with this view" means that many people _____ that brain drain harms the economies of poor countries.

- a) support the idea
- b) disagree with the idea
- c) raise the issue
- d) are aware

Doğru cevap: b

Yandaki maddenin **erkekler** lehine işlediği yapılan analiz sonucu ortaya konmuştur. Bu maddenin yanlı çıkmasına katılma durumunuzu belirtiniz.

Kesinlikle Katılıyorum ()

Katılıyorum ()

Katılmıyorum(X)

Cevabınızın gerekçesini belirtiniz.

Soruda cinsiyet açısından yanlılığa neden olacak bir ipucu veya kelime dikkatim çekmedi.

ÖZGEÇMİŞ

Adı Soyadı	Gülşen TAŞDELEN TEKER
Doğum Yeri	Ankara
Doğum Yılı	1984
Medeni Hali	Evli

Eğitim ve Akademik Durumu

Lise	Dikmen Yabancı Dil Ağırlıklı Lisesi, Ankara	1998-2002
Lisans	Orta Doğu Teknik Üniversitesi Eğitim Fakültesi İlköğretim Fen Bilgisi Öğretmenliği	2002-2007
Yüksek Lisans	Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü Eğitim Bilimleri Anabilim Dalı Eğitimde Ölçme ve Değerlendirme Bilim Dalı	2007-2009
Doktora	Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Eğitim Bilimleri Anabilim Dalı Eğitimde Ölçme ve Değerlendirme Bilim Dalı	2010-2014
Yabancı Dil	İngilizce	KPDS 91
İş Deneyimi	Sakarya Üniversitesi Hendek Eğitim Fakültesi Eğitim Bilimleri Bölümü Eğitimde Ölçme ve Değerlendirme Anabilim Dalı Görev: Öğretim Görevlisi	2010- (devam ediyor)
	Bursa İnegöl Altıeylül İlköğretim Okulu Görev: Fen ve Teknoloji Öğretmeni	2009-2010