

**TÜRKİYE CUMHURİYETİ  
ANKARA ÜNİVERSİTESİ  
SAĞLIK BİLİMLERİ ENSTİTÜSÜ**

**KAS-İSKELET SİSTEMİ SORUNLARININ ÖZÜRLÜLÜK  
DEĞERLENDİRİMİNDE BİLGİSAYAR UYARLAMALI TEST  
YÖNTEMİNİN UYGULANMASI**

**Derya ÖZTUNA**

**BİYOİSTATİSTİK ANABİLİM DALI  
DOKTORA TEZİ**

**DANIŞMAN**

**Doç.Dr. Atilla Halil ELHAN**

Bu tez, Ankara Üniversitesi Bilimsel Araştırma Projeleri Müdürlüğü tarafından  
0809241 proje numarası ile desteklenmiştir.

**2008-ANKARA**

# İÇİNDEKİLER

Kabul ve Onay	ii
İçindekiler	iii
Önsöz	vi
Simgeler ve Kısaltmalar	vii
Şekiller	ix
Çizelgeler	xi
<b>1. GİRİŞ</b>	<b>1</b>
1.1. Araştırmanın Konusu, Amacı ve Önemi	1
1.2. Klasik Test Teorisi ve Madde Yanıt Teorisinin Tarihçesi	6
1.3. Klasik Test Teorisi (Classical Test Theory)	8
1.3.1. Klasik Test Teorisinin Temel Özellikleri	8
1.3.2. Klasik Test Teorisinin Varsayımları	8
1.3.3. Klasik Test Teorisinde Madde ve Ölçme Aracı İstatistikleri	9
1.3.4. Klasik Test Teorisinin Avantajları	9
1.3.5. Klasik Test Teorisinin Dezavantajları	9
1.4. Madde Yanıt Teorisi (Item Response Theory)	10
1.4.1. Madde Yanıt Teorisinin Temel Özellikleri	10
1.4.2. Madde Yanıt Teorisinin Varsayımları	11
1.4.3. Madde Yanıt Teorisinde Madde ve Ölçme Aracı İstatistikleri	13
1.4.3.1. Madde Parametreleri	13
1.4.3.2. Madde ve Ölçme Aracı Bilgisi (Item and Test Information)	15
1.4.4. Madde Yanıt Teorisinde Kullanılan Modeller	16
1.4.4.1. İki Sonuçlu Puanlama Modelleri	18
1.4.4.2. Çok Sonuçlu Puanlama Modelleri	19
1.4.5. İncelenen Özelliğin ve Madde Parametrelerinin Tahmini	22
1.4.6. Model-Veri Uyumunun ve Madde İşlev Farklılığının Belirlenmesi	23
1.4.6.1. Model-Veri Uyumunun Belirlenmesi	23
1.4.6.2. Madde İşlev Farklılığının Belirlenmesi	25
1.4.7. Madde Yanıt Teorisinin Avantajları	26
1.4.8. Madde Yanıt Teorisinin Dezavantajları	27
1.4.9. Ölçme Aracı Geliştirme Açısından İki Teorinin Karşılaştırılması	27

1.5. Bilgisayar Uyarlamalı Test Yöntemi	29
1.5.1. Bilgisayar Uyarlamalı Test Yöntemine Giriş	29
1.5.2. Bilgisayar Uyarlamalı Test Yönteminin Tarihsel Gelişimi	30
1.5.3. Bilgisayar Uyarlamalı Test Yönteminin İşleyişi	31
1.5.3.1. Bilgisayar Uyarlamalı Test Yönteminin İki Sonuçlu Maddelerde İşleyişi	34
1.5.3.2. Bilgisayar Uyarlamalı Test Yönteminin Çok Sonuçlu Maddelerde İşleyişi	35
1.5.4. Bilgisayar Uyarlamalı Test Yönteminde Önemli Noktalar	38
1.5.4.1. İlk Maddenin Seçimi	38
1.5.4.2. Madde Seçimi	38
1.5.4.3. Durdurma Kuralları	39
1.5.4.4. Soru Bankası	39
1.5.5. Bilgisayar Uyarlamalı Test Yönteminin Avantajları	40
1.5.6. Bilgisayar Uyarlamalı Test Yönteminin Dezavantajları	41
<b>2. GEREÇ VE YÖNTEM</b>	42
<b>3. BULGULAR</b>	56
3.1. Bel Ağırlı Hastalar İçin Elde Edilen Bulgular	56
3.1.1. Demografik Bulgular	56
3.1.2. Başlangıç Soru Bankasının Klasik Yöntemle Uygulanması	56
3.1.3. Açıklayıcı ve Doğrulayıcı Faktör Analizi Sonuçları	57
3.1.4. Rasch Analizi Sonuçları	60
3.1.4.1. “Ağrı-Uyku-Duygusal Durum” Soru Bankasının Rasch Analizi	60
3.1.4.2. “Aktivite-Katılım” Soru Bankasının Rasch Analizi	65
3.1.5. Benzetim Çalışması ve Gerçek Uygulama Sonuçları	70
3.1.5.1. “Ağrı-Uyku-Duygusal Durum” Soru Bankası için Benzetim Çalışması ve Gerçek Uygulama Sonuçları	71
3.1.5.2. “Aktivite-Katılım” Soru Bankası için Benzetim Çalışması ve Gerçek Uygulama Sonuçları	73
3.2. Diz Osteoartritli Hastalar İçin Elde Edilen Bulgular	76
3.2.1. Demografik Bulgular	76
3.2.2. Başlangıç Soru Bankasının Klasik Yöntemle Uygulanması	76
3.2.3. Açıklayıcı ve Doğrulayıcı Faktör Analizi Sonuçları	77
3.2.4. Rasch Analizi Sonuçları	81
3.2.4.1. “Aktivite-Ağrı” Soru Bankasının Rasch Analizi	81
3.2.4.2. “Ağrı-Sosyal-Bilişsel-Duygusal Durum” Soru Bankasının Rasch Analizi	87

3.2.5. Benzetim Çalışması ve Gerçek Uygulama Sonuçları	92
3.2.5.1. “Aktivite-Ağrı” Soru Bankası için Benzetim Çalışması ve Gerçek Uygulama Sonuçları	93
3.2.5.2. “Ağrı-Sosyal-Bilişsel-Duygusal Durum” Soru Bankası için Benzetim Çalışması ve Gerçek Uygulama Sonuçları	96
<b>4. TARTIŞMA</b>	99
<b>5. SONUÇ VE ÖNERİLER</b>	106
<b>ÖZET</b>	109
<b>SUMMARY</b>	110
<b>KAYNAKLAR</b>	111
<b>ÖZGEÇMİŞ</b>	118

## ÖNSÖZ

Kas-İskelet Sistemi Sorunlarının Özürllük Deęerlendiriminde Bilgisayar Uyarlamalı Test Yönteminin Uygulanması amacıyla gerçekleştirilen bu çalışmanın yürütülmesinde, desteęini esirgemeyen, çalışmamı görüş ve önerileriyle yönlendiren deęerli danışmanım Doę.Dr. Atilla Halil ELHAN'a şükranlarımı sunmayı borç bilirim.

Çalışma konusunun belirlenmesinde sunduęu önerilerden dolayı Prof. Dr. Alan TENNANT'a, çalışma boyunca konu ile ilgili olarak bilgilerini ve görüşlerini her zaman paylaşan ve desteklerini her zaman hissettiğim Prof. Dr. Ayşe A. KÜÇÜKDEVECİ ve Prof.Dr. Şehim KUTLAY'a teşekkürlerimi ve saygılarımı sunarım.

Veri toplama aşamasında gösterdikleri ilgi ve yardımdan dolayı Fiziksel Tıp ve Rehabilitasyon Anabilim Dalı araştırma görevlilerine ve Biyoistatistik Anabilim Dalında görev yapan deęerli çalışma arkadaşlarıma teşekkürlerimi sunarım.

Çalışma kapsamında yazılan bilgisayar programını C programlama dili ile yazarak gerçek uygulama öncesinde yapılan pilot çalışma için gerekli alt yapıyı hazırlayan Murat ÖZKUT'a teşekkürlerimi ve sevgilerimi sunarım.

Tez çalışmasının uygulanması için gerekli bilgisayar programının yazılmasında ve sonuçların elde edilmesinde emeęi geçen ve manevi desteęini her zaman yanımda hissettiğim eşim, Bilgisayar Mühendisi A. Emre ÖZTUNA'ya teşekkürlerimi sunarım.

Ayrıca çalışma sırasında karşılaştığım zorlukların üstesinden gelmemde manevi desteklerini hiçbir zaman esirgemeyen ve yanımda olan aileme özel teşekkürlerimi sunarım.

## SİMGELER VE KISALTMALAR

a	Madde Ayırt Edicilik Parametresi
ACT	American College Test
AFA	Açıklayıcı Faktör Analizi
AKOK	Artık Kareler Ortalaması Karekökü
AOAK	Ağırlıklandırılmış Ortalama Artıkların Karekökü
ASVAB	Armed Services Vocational Aptitude Battery
b	Madde Zorluk Parametresi
BAİ	Birey Ayırsama İndeksi
BUT	Bilgisayar uyarlamalı test
c	Şansa Bağlı Tahmin Parametresi
DFA	Doğrulayıcı Faktör Analizi
DSÖ	Dünya Sağlık Örgütü
DUIİ	Düzeltilmiş Uyum İyiliği İndeksi
E	Hata Puanı
ETS	The Educational Testing Service
FTR	Fiziksel Tıp ve Rehabilitasyon
GMAT	Graduate Management Admission Test
GRE	The Graduate Record Examination
HAQ	Health Assessment Questionnaire
ICC	Intraclass Correlation Coefficient
ICF	International Classification of Functioning, Disability and Health
KTT	Klasik Test Teorisi
KUIİ	Karşılaştırmalı Uyum İndeksi
L	Olabilirlik Fonksiyonu
MİF	Madde İşlev Farklılığı
MKF	Madde Karakteristik Fonksiyonu
MYT	Madde Yanıt Teorisi
NHP	Nottingham Health Profile
$P_i(\theta)$	i. Maddeye Doğru Yanıt Verme Olasılığı
$\theta$	İncelenen Özellik
SAT	Scholastic Aptitude Tests and Scholastic Assessment Tests

SKK	Sınıfıçı Korelasyon Katsayısı
T	Gerçek Puan
TBA	Temel Bileşenler Analizi
TLİ	Tucker-Lewis İndeksi
TOEFL	The Test of English as a Foreign Language
Uİİ	Uyum İyiliği İndeksi
WHO	World Health Organization
WHODAS-II	World Health Organization-Disability Assessment Schedule II
WOMAC	Western Ontario and McMaster Universities
X	Gözlenen Puan
YEM	Yapısal Eşitlik Modelleri
YHKOK	Yaklaşımın Hata Kareler Ortalaması Karekökü
$I_i(\theta)$	i. Madde İçin Bilgi Fonksiyonu
$I(\theta)$	Ölçme Aracı Bilgi Fonksiyonu
$SE(\theta)$	Tahminin Standart Hatası

## ŞEKİLLER

Şekil 1.1: Ayırt edicilik parametreleri aynı, zorluk parametreleri farklı olan üç maddeye ilişkin MKF eğrileri	14
Şekil 1.2: Zorluk parametreleri aynı, ayırt edicilik parametreleri farklı olan üç maddeye ilişkin MKF eğrileri	15
Şekil 1.3. İki Sonuçlu Puanlamaya Sahip Bir Madde için MKF	18
Şekil 1.4. Çok Sonuçlu Puanlamaya Sahip Bir Madde için MKF	20
Şekil 1.5: BUT Yönteminin Mantığı	32
Şekil 1.6: BUT Yönteminin İki Sonuçlu Maddelerde İşleyişi	34
Şekil 1.7: BUT Yönteminin Çok Sonuçlu Maddelerde İşleyişi	36
Şekil 2.1: Eşik Değerlerinin Sıralı Olduğu Örnek Bir Madde	50
Şekil 2.2: Eşik Değerlerinin Sıra Takip Etmediği Örnek Bir Madde	50
Şekil 3.1. “Ağrı-Uyku-Duygusal Durum” Soru Bankasındaki Maddelerin Eşik Değerleri	61
Şekil 3.2. “Ağrı-Uyku-Duygusal Durum” Final Soru Bankasındaki Maddelerin ve Hastaların $\theta$ Düzeyi Boyunca Dağılımı	64
Şekil 3.3. “Aktivite-Katılım” Soru Bankasındaki Maddelerin Eşik Değerleri	66
Şekil 3.4. “rm9” Kodlu Maddenin Yaş Gruplarına Göre MKF Eğrileri	67
Şekil 3.5. “Aktivite-Katılım” Final Soru Bankasındaki Maddelerin ve Hastaların $\theta$ Düzeyi Boyunca Dağılımı	69
Şekil 3.6. Gerçek Uygulama Sonuçlarına Göre “Ağrı-Uyku-Duygusal Durum” Final Soru Bankası İçin Uyum Sınırlarına İlişkin Blant-Altman Grafiği	73
Şekil 3.7. Gerçek Uygulama Sonuçlarına Göre “Aktivite-Katılım” Final Soru Bankası İçin Uyum Sınırlarına İlişkin Blant-Altman Grafiği	75
Şekil 3.8. “Aktivite-Ağrı” Soru Bankasındaki Maddelerin Eşik Değerleri	81
Şekil 3.9. “wd55” Kodlu Maddenin Cinsiyete Göre MKF Eğrileri	84



Şekil 3.10. “Aktivite-Ağrı” Final Soru Bankasındaki Maddelerin ve Hastaların $\theta$ Düzeyi Boyunca Dağılımı	87
Şekil 3.11. “Ağrı-Sosyal-Bilişsel-Duygusal Durum” Soru Bankasındaki Maddelerin Eşik Değerleri	88
Şekil 3.12. “n26” Kodlu Maddenin Cinsiyete Göre MKF Eğrileri	89
Şekil 3.13. “Ağrı-Sosyal-Bilişsel-Duygusal Durum” Final Soru Bankasındaki Maddelerin ve Hastaların $\theta$ Düzeyi Boyunca Dağılımı	92
Şekil 3.14. Gerçek Uygulama Sonuçlarına Göre “Aktivite-Ağrı” Final Soru Bankası İçin Uyum Sınırlarına İlişkin Blant-Altman Grafiği	95
Şekil 3.15. Gerçek Uygulama Sonuçlarına Göre “Ağrı-Sosyal-Bilişsel-Duygusal Durum” Final Soru Bankası İçin Uyum Sınırlarına İlişkin Blant-Altman Grafiği	98

## ÇİZELGELER

Çizelge 1.1. MYT'nin Tarihçesi (1982 yılına kadar)	7
Çizelge 1.2. Puanlama Formatına Göre MYT Modelleri	17
Çizelge 3.1. Bel Ağrılı Hastalar İçin "Ağrı-Uyku-Duygusal Durum" Soru Bankası	58
Çizelge 3.2. Bel Ağrılı Hastalar İçin "Aktivite-Katılım" Soru Bankası	59
Çizelge 3.3. "Ağrı-Uyku-Duygusal Durum" Final Soru Bankasındaki Maddelerin Uyum İstatistikleri	63
Çizelge 3.4. "Aktivite-Katılım" Final Soru Bankasındaki Maddelerin Uyum İstatistikleri	68
Çizelge 3.5. Diz Osteoartritli Hastalar İçin "Aktivite-Ağrı" Soru Bankası	78
Çizelge 3.6. Diz Osteoartritli Hastalar İçin "Ağrı-Sosyal-Bilişsel-Duygusal Durum" Soru Bankası	80
Çizelge 3.7. "Aktivite-Ağrı" Final Soru Bankasındaki Maddelerin Model Uyum İstatistikleri	85
Çizelge 3.8. "Ağrı-Sosyal-Bilişsel-Duygusal Durum" Final Soru Bankasındaki Maddelerin Model Uyum İstatistikleri	90

# 1. GİRİŞ

## 1.1. Araştırmanın Konusu, Amacı ve Önemi

Kas-iskelet sistemi hastalıkları, toplumda prevalansı en yüksek hastalık gruplarından biri olup; toplumdaki ekonomik yükü de oldukça fazladır. İngiltere, Amerika Birleşik Devletleri ve Kanada gibi batı ülkelerindeki istatistiksel veriler, bu hastalıklara bağlı harcamaların milli gelirin %1-2.5'una eriştiğini göstermektedir. Kas-iskelet sistemi hastalıklarının yol açtığı fiziksel ve fonksiyonel bozukluklardan dolayı kişilerin yaşam kalitesindeki olumsuz etkilenmenin diğer kronik hastalıklara göre daha fazla olduğu gösterilmiştir (Küçükdeveci, 2004). Yine kronik bel ağrıları genel popülasyonun %75'ini etkileyen bir sorun olup, yüksek oranda özürlülük nedeni olabilmekte ve işgücü kaybına yol açmaktadır (Suarez-Almazor ve ark., 2000). Bu hastalıklar, kronik sorunlar olduklarından hastaların uygun şekilde izlemleri gerekmekte ve bu izlemlerde objektif parametrelerin ve standart değerlendirme prosedürlerinin kullanılması önerilmektedir.

Birçok sağlık alanında olduğu gibi fiziksel kas-iskelet sorunlarında hasta değerlendirmesinde Dünya Sağlık Örgütü (DSÖ) [World Health Organization-WHO]'nın "Uluslararası Bozukluk, Disabilite ve Handikap Sınıflaması" temel yapısal model olarak alınmaktadır. Bu model, 2001'de revize edilerek "Uluslararası Fonksiyon, Disabilite ve Sağlık" [International Classification of Functioning, Disability and Health-ICF] olarak yayınlanmıştır (WHO, 2001). ICF, çeşitli disiplinler ve farklı sektörlere hizmet etmek amacıyla tasarlanmış kapsamlı bir sınıflandırma olup; özel amaçları şunlardır: 1) Sağlık ve sağlıkla ilgili durumların, sonuçlarının ve belirleyicilerinin anlaşılması ve araştırılması için bilimsel bir temel oluşturmak, 2) Sağlık çalışanları, araştırmacılar, siyasetçiler gibi ve özürlülüğü olanlar da dahil olmak üzere toplumda farklı kullanıcılar arasında sağlık ve sağlıkla ilgili durumlarla ilgili iletişimi arttırmak

amacıyla ortak bir dil oluşturmak, 3) Ülkeler, sağlıkla ilgili disiplinler, hizmetler ve zaman açısından verilerin karşılaştırılmasına olanak sağlamak, 4) Sağlıkla ilgili bilgi sistemleri için sistematik kodlama şemaları sağlamak.

Revize edilen son sınıflamaya göre vücut fonksiyonları (*body functions*), vücut sistemlerinin fizyolojik fonksiyonları, vücut yapıları (*body structures*) ise vücudun anatomik bölümleridir. Bozukluklar (*impairments*), vücut yapı ve fonksiyonlarındaki anlamlı sapma ya da kayıp gibi sorunlardır. Aktivite (*activity*), birey tarafından bir hareket ya da görevin yerine getirilmesidir. Katılım (*participation*), bir yaşam durumuna yani sosyal hayata iştirak etmeyi ifade etmektedir. Aktivite sınırlanması (*activity limitation*), kişinin aktivitelerini yerine getirmedeki zorluklardır. Katılımın kısıtlanması (*participation restriction*), kişinin yaşam durumlarına yani sosyal hayata iştirak etmesindeki sorunlardır. Çevresel faktörler (*environmental factors*) ise kişinin yaşamını sürdürdüğü ortamdaki fiziksel ve sosyal çevredir. Bu yeni sınıflamada fonksiyon görme (*functioning*), vücut fonksiyonları/yapıları, aktivite ve katılımı içeren bir şemsiye terim; özürülük (*disability*) ise bozukluklar, aktivite sınırlanması ve katılımın kısıtlanmasını içine alan bir şemsiye terim olarak belirtilmiştir.

DSÖ, bu model içine “yaşam kalitesi” terimini dahil etmemiştir. Yaşam kalitesi, “sübjektif iyilik hali” veya bir diğer ifadeyle “kişinin kendi yaşamından memnun olma durumu” olarak tanımlanmaktadır (Fuhrer, 1994; Whalley ve ark., 1997). DSÖ ise benzer şekilde yaşam kalitesini “Bireyin, gerek kültürel ve içinde bulunduğu ortamın değer yargıları, gerekse kendi hedefleri, beklentileri, standartları ve ilgileri bağlamında, hayatta kendi durumunu algılama biçimi” olarak tanımlamıştır.

Tıp alanında sonuç, “Bir ölçüm ya da durumdaki beklenen ya da aranan değişme” olarak tanımlanmaktadır. Sonuç ölçümü, “hasta izleminde karar verme”, “hastalık sonuçlarını değerlendirme”, “sağlık politikaları belirleme”, “farmasötik endüstriyel alandaki araştırma, geliştirme ve değerlendirmenin daha iyi yapılmasını sağlama” amaçlarıyla yapılmaktadır. Sonuç ölçümü amacıyla kullanılan parametreler bozukluk, aktivite, katılım ve/veya yaşam kalitesi düzeyinde değerlendirim yapabilirler (Wade, 2003).

Kas-iskelet sistemi sorunlarında kullanılan ölçme araçları gerek değerlendirdiği alanlar (bozukluk, aktivite, yaşam kalitesi gibi), gerekse aynı alanların farklı ölçme araçları ile değerlendirimi nedeniyle çeşitlilik göstermektedir. Bu ölçme araçlarından her birinin kendine ait uygulama şekli, avantaj ve dezavantajları vardır. Sık kullanılan ölçme araçlarının genellikle kullanıldıkları toplumlar için geçerlilik ve güvenilirlik çalışmaları yapılmıştır (Beaton ve ark., 2000; Guillemin, 2000). Kas-iskelet sistemi sorunlarında sonuç değerlendirmede kullanılan ölçme araçlarının bu denli çeşitlilik göstermesi kullanıcıların ortak bir dilde konuşmasını engelleyerek, hasta değerlendirim ve izleminde standardizasyonu olumsuz etkilemektedir. Dolayısıyla sağlık çalışanları arasında ortak bir dil kullanımının sağlanması son derece önemlidir (Tennant ve ark., 2004). Günümüzde, uygulamada ICF sınıflamasında tanımlandığı şekilde kas-iskelet sistemi sorunlarında “özürlülük” kavramını değerlendiren ideal bir ölçme aracı (soru bankası/ölçek) bulunmamaktadır. Kas-iskelet sistemi sorunlarında, özürlülük düzeyinin belirlenmesi, “bozukluk”, “aktivite sınırlanması”, “katılım kısıtlanması” ve “yaşam kalitesi”ni değerlendiren maddeleri içeren ideal bir ölçme aracı ile yapılmalıdır.

Günlük yaşantımızda ve bilimsel çalışmalarda ölçme önemli bir yer tutar. Ölçme, genel anlamıyla herhangi bir niteliği gözlemek ve gözlem sonuçlarını sayılarla veya başka sembollerle ifade etmektir (Turgut, 1977). Bir çocuğun boyunun/kilosunun ölçülmesi “doğrudan ölçme” kapsamına girerken, aynı çocuğun herhangi bir dersteki yetenek seviyesinin ölçülmesi

“dolaylı ölçme” kapsamına girmektedir. Doğrudan ölçmede, bir nesneyi/bireyi bir kez ölçmek yeterli olabilirken; dolaylı ölçmede tekrarlanan her bir ölçümde farklı sonuçlar elde edilebilmektedir. Çünkü dolaylı ölçmede, doğrudan ölçme için söz konusu olan hata kaynaklarından (ölçücü, ölçme aracı ve ölçülen özellik) başka, asıl ölçme konusu olan davranış ile gerçekte gözlenen davranış arasında kurulan ilişki de bir hata kaynağı olabilir (Tekin, 2002).

Doğrudan gözlenemeyen (gizli/latent) değişkenleri ölçmede; hatayı en aza indirmek, ölçmek istediğimiz niteliğe başka değişkenleri karıştırmamak ve daha duyarlı ölçme sonuçları elde etmek, diğer bilim dallarında olduğu gibi sağlık alanındaki sonuç ölçümlerinde de geçerlidir. Ölçme araçları, ölçülen büyüklüğün sayılarla ifade edilmesini; böylelikle gözlemin daha duyarlı yapılarak objektif ve standart ölçümler elde edilmesini sağlar. Ölçme araçlarından elde edilen sonuçlar bireylerin incelenen özellik ( $\theta$ ) düzeyleri hakkında değer yargılarına varmada kullanılır. Varılan değer yargılarının doğruluğu, ölçme araç veya yöntemlerinin niteliklerine bağlıdır (Anıl, 2002).

Bir ölçme aracının sahip olması istenen başlıca dört özelliği vardır. Bunlardan ilki geçerlilik (validity) olup; ölçme aracının  $\theta$  düzeyini, tam ve doğru olarak, başka özelliklerle karıştırmadan ölçebilmesidir. İkinci özellik olan güvenilirlik (reliability) ise; ölçme aracının  $\theta$  düzeyini değişik zamanlarda benzer olarak ölçebilmesidir ki, bu özellik ölçeğin tutarlılığını da gösterir.  $\theta$  düzeyini doğru ve tutarlı bir biçimde ölçen bir aracın geliştirilmesinin, uygulanmasının ve değerlendirilmesinin kolay ve düşük maliyetli olması da istenir. Bu üçüncü özelliğe kullanılabilirlik (practicability/usefulness) adı verilir (Alpar, 2003). Dördüncüsü, ölçme aracının, yapılan bir müdahale sonrasında, klinik olarak anlamlı bir farkı belirleyebilme özelliği olup; yanıt verme (responsiveness) olarak adlandırılır ve etki büyüklüğü ile değerlendirilir (Kaziz ve ark., 1989).

Uygun maddeler seçerek istenen nitelikte bir ölçme aracı geliştirme sürecinde madde analiz çalışmaları önemli bir yer tutmaktadır. Bu amaçla kullanılan iki teori, klasik test teorisi (KTT) ve madde yanıt teorisidir (MYT). Her yıl, daha güvenilir, geçerli, duyarlı ve yorumlanabilir ölçme araçları elde etme ümidiyle, sağlık alanında yeni sonuç ölçümleri geliştirilmekte veya mevcut olanlar tekrar gözden geçirilmektedir. Geçtiğimiz on yılda, madde ve ölçme aracının analizinde, ölçme araçlarının puanlanmasında ve bilgisayar uyarlamalı test (BUT) yönteminde sağladığı yararlar nedeniyle MYT'nin sağlık araştırmalarında kullanımı dikkate değer bir şekilde artmıştır.

Bu çalışmanın amacı, sık görülen kas-iskelet sistemi sorunlarından bel ağrısı ve diz osteoartritinde özürülük değerlendiriminde kullanılacak, ICF alanlarını temel alan birer ölçme aracı geliştirmek ve bu ölçme araçlarında yer alan maddelerle BUT yöntemini uygulayarak hastaların özürülük düzeylerini daha kısa sürede ve daha az madde ile saptamaya çalışmaktır.

Dünya literatüründe bel ağrılı ve diz osteoartritli hastaların özürülük durumlarını belirleyen, ICF alanlarını temel alan spesifik ölçme araçları bulunmamaktadır. Bu bağlamda bu iki soruna spesifik olarak geliştirilecek olan ölçme araçları, bu hastalarda özürülük değerlendirmenin standardizasyonu açısından önemlidir.

BUT yöntemi yurtdışında özellikle eğitim alanında TOEFL (*The Test of English as a Foreign Language*), GRE (*The Graduate Record Examination*) ve GMAT (*Graduate Management Admission Test*) gibi bazı sınavların değerlendirilmesinde rutin olarak uygulanmaktadır. Uluslararası alanda sağlık alanında kullanımı henüz araştırma düzeyinde olup; çok iyi sonuçlar vermiştir. Bu yöntem ülkemizde eğitim alanında araştırma düzeyinde kullanılmış olup, sağlık alanında uygulaması bulunmamaktadır. Dolayısıyla bu çalışma ülkemizde sağlık alanında BUT yönteminin kullanılması bakımından benzer çalışmalar için öncü olacaktır.

## 1.2. Klasik Test Teorisi ve Madde Yanıt Teorisinin Tarihçesi

Hulin ve ark. (1983) dolaylı ölçmeyi; bireylerin/nesnelerin özelliklerine belirli kurallara uyarak sayısal değerler verme olarak tanımlamaktadır. Psikometrik teoriler, bireylerin belirli bir psikolojik ölçme aracına gösterdikleri tepkileri, gözlenen bu tepkilerin altında yatan özellikleri tahmin etmede uyulacak kuralları belirlemekle ilgilenir.

Francis Galton'un, 19. yy'da çeşitli alanlarda gösterilen performansla, bireysel farklılıklar arasında ilişki kurması ile bireysel farklılıkların ortaya çıkması, ölçme aracı geliştirme çalışmalarının başlangıcı olarak değerlendirilmiştir. Yeteneğin ( $\theta$ 'nın) değerlendirilmesinde temel adımlar 20.yy'ın ilk çeyreğinde Binet, Terman ve Cattell'in yanı sıra; Thorndike ve Thurstone tarafından atılmıştır. Gulliksen, 1958 yılında grup değişikliklerinin madde parametreleri üzerine etkisini araştıran deneysel ve kuramsal çalışmaların çok yetersiz olduğuna değinmiş, test edilen gruptaki sistematik değişikliklerde, madde parametrelerinin görelilik olarak değişmez kalmasını sağlayacak deneysel ve kuramsal çalışmaların gereğine işaret etmiştir.

Ölçme aracı geliştirmede bu eksikliğe cevap veren teoriler, Lord tarafından geliştirilen MYT'dir. Hambleton ve Swaminathan, 1970'lerde ölçme uzmanlarının çalışmalarında en yaygın konu haline gelen MYT'nin temellerinin aslında, 1930 ve 1940'lı yıllarda atıldığına işaret etmiştir. Ayrıca, MYT'nin temel kavramı olan "madde karakteristik fonksiyonu (MKF)"nın 1946'larda Leylard Tucker'a atfedilebileceğini belirtmişlerdir.

Birnbaum, Lord ve diğerlerinin 1952 yılından itibaren yaptıkları çalışmalarda, madde analizleri için formüle ettikleri modellerin hepsi, her madde için bir, iki ya da üç parametrenin tahminini gerektirmektedir. Bu parametreler; madde zorluk parametresi, madde ayırt edicilik parametresi ve şansa bağlı tahmin parametresidir.



Danimarkalı matematikçi Georg Rasch, basit bir lojistik model geliştirerek, ölçme aracı geliştirme çalışmalarına önemli katkıda bulunmuştur. Rasch, 1945'lerden itibaren başladığı ölçme aracı geliştirme çalışmaları ile tek parametrelili lojistik modeli önermiş; 1960 yılında madde zorluk parametresini, üzerinde çalışılan örneklemden bağımsız olarak tahmin etmiştir. Günümüzde bu model, birçok bilim adamı tarafından ölçme aracı geliştirmede standart olarak kullanılmaktadır. MYT'nin tarihçesi Çizelge 1.1'de sunulmuştur (Hambleton ve Swaminathan, 1985; Reeve, 2002).

Çizelge 1.1. MYT'nin Tarihçesi (1982 yılına kadar)

Yıl	Kişi	Olay
1916	Binet & Simon	Performans düzeylerine karşı bir bağımsız değişkenin grafiğini çizerek, bu grafiği ölçme aracı geliştirmede kullanmışlardır.
1925	Thurstone	MYT'nin kavramsal oluşumuna "A Method of Scaling Psychological and Educational Tests" adlı makalesinde yer vermiştir.
1936	Richardson	MYT model parametreleri ile klasik madde parametreleri arasındaki ilişkiyi ortaya çıkarmıştır.
1943	Richardson ve Ferguson	Maddeler için doğru yanıt verme oranlarını, normalleştirilmiş skorların bir fonksiyonu olarak göstermek için normal ogive modelini ortaya koymuşlardır.
1943-44	Lawley	Normal ogive eğrisinin özelliklerinin istatistiksel analizini geliştirmiş ve madde parametreleri ile bu parametrelerin doğrusal yaklaşımları için en çok olabilirlik tahmin yöntemlerini tanımlamıştır.
1950	Lazarsfeld	Madde yanıtları arasında gözlenen ilişkileri hesaba katan gözlenmeyen bir değişken tanımlamıştır.
1952	Lord	Gizli yapı fikrini ortaya atmıştır ve bu yapıyı gözlenen test skorundan ayırmıştır. İki parametrelili normal ogive modelini tanımlamış, model parametre tahminlerini ortaya çıkarmış, modelin uygulamalarını değerlendirmiştir.
1957-58	Birnbaum	Normal ogive modeli yerine daha kolay lojistik modeller kullanmıştır.
1960	Rasch	3 farklı madde yanıt modeli geliştirmiş ve bunları " <i>Probabilistic Models For Some Intelligence And Attainment Test</i> " adlı kitapta tanımlamıştır.
1967	Wright	Rasch modeli araştırmalarının 1970'li yıllarda Amerika'daki lideridir. "ETS Invitational Conference on Testing Problems"daki sunumu, özellikle Rasch modeli ile MYT çalışmalarına temel olmuştur. Daha sonra, kendisinin geliştirdiği oldukça başarılı olan AERA Rasch modeli deneme programları, Rasch modelinin araştırmacılar tarafından anlaşılmasına önemli katkıda bulunmuştur.
1968	Lord & Novick Birnbaum	Lord ve Novick'in "Statistical Theories of Mental Test Scores" adlı kitabı, psikometri alanında önemli bir aşama olup; KTT'nin istatistiksel değerlendirmesi hakkında bilgi verir. Allen Birnbaum tarafından yazılan kitabın diğer yarisında, MYT modellerinin tanımına yer verilmiştir.
1969	Wright & Panchapakesan	BICAL programını geliştirmişlerdir. Bu program, Rasch modeli uygulamalarına getirdiği kolaylıklar nedeniyle çok önemli bir yere sahiptir.
1969	Samejima	Geliştirdiği model, hem çok kategorili hem de sürekli yanıtlar için kullanılmıştır. Tek boyutlu modelleri, çok boyutlu modellere genişletmiştir.
1972	Bock	Parametre tahmini hakkında önemli fikirler vermiştir.
1974	Fischer	Doğrusal lojistik modelleri ile ilgili geniş araştırma programını tanımlamıştır.
1976	Lord	Lojistik test modeli ile parametre tahmini yapmak için kullanılan LOGIST bilgisayar programını kullanılabilir hale getirmiştir.
1977	Baker	Parametre tahmin yöntemlerini geniş bir şekilde yeniden incelemiştir.
1979	Wright & Stone	" <i>Best Test Design</i> " adlı kitapta, Rasch modelinin temelindeki teoriyi ve uygulamaları tanımlamışlardır.

### **1.3. Klasik Test Teorisi (Classical Test Theory)**

#### **1.3.1. Klasik Test Teorisinin Temel Özellikleri**

KTT; temelleri 1905 yılında Spearman tarafından atılan, gerçek puanın, ölçme sonucunda elde edilen gözlenen puandan tahmin edilmeye çalışıldığı bir teoridir. Çünkü gerçek değer, ölçmek istediğimiz değişkenin başka değişkenlerle karışması, ölçme aracının hatalı olması, ölçücünün hata yapması gibi nedenlerden dolayı doğrudan elde edilemez.

Bireyin  $\theta$  düzeyinin; ölçmenin yapıldığı anda sahip olduğu değeri, onun gerçek puanı olup; ölçme süresince değişmez. Ölçme sonucunda elde edilen değer de bireyin gözlenen puanıdır. KTT'de bir bireyin bir ölçme aracına ilişkin gözlenen puanı (X); gerçek puan (T) ve hata puanı (E) toplamından oluşur ve  $X=T+E$  eşitliği, KTT'nin temel denklemini oluşturur (Lord, 1980). KTT'de bireyin  $\theta$  düzeyi, ölçme aracının maddelerine verdiği yanıtların toplamıdır (ham puan).

#### **1.3.2. Klasik Test Teorisinin Varsayımları**

KTT aşağıdaki varsayımlara dayandırılmıştır (Doğan, 2002):

1. E değerlerinin beklenen değeri sıfırdır.
2. T ile E değerleri arasında ilişki yoktur.
3. İki ayrı ölçmeye ait E değerleri arasında ilişki yoktur.
4. İki ayrı ölçmeye ait E ve X değerleri arasında ilişki yoktur.

### **1.3.3. Klasik Test Teorisinde Madde ve Ölçme Aracı İstatistikleri**

KTT'de madde istatistikleri, madde zorluk parametresi ve madde ayırt edicilik parametresi; ölçme aracı istatistikleri ise ham puan ile güvenilirlik ve geçerlilik katsayılarıdır. Madde zorluk parametresi, bir maddeye doğru yanıt verenlerin sayısının, o maddeyi yanıtlayanların sayısına oranı olarak tanımlanır. Madde ayırt edicilik parametresi ise, bir maddenin, o maddeyle ölçülmek istenen özelliğe sahip olanla olmayanı ayırt edebilmesi olarak tanımlanır. Değişkenlerin ikisi de sürekli kabul edildiğinde Pearson ilişki katsayısıyla, değişkenler kesikli/yapay kesikli kabul edildiğinde ise çift serili/nokta çift serili ilişki katsayısıyla hesaplanır (Alpar, 2003).

### **1.3.4. Klasik Test Teorisinin Avantajları**

KTT, geliştirilen ilk modern ölçme teorisi olma özelliği taşımaktadır. KTT'nin en büyük avantajı, teorinin varsayımlarının çeşitli uygulamalarda kolaylıkla test edilebilmesidir. Bu teori, uygulama ve parametreleri tahmin etmedeki kolaylıklar nedeniyle diğer teorilerden daha yaygın olarak kullanılmaktadır.

### **1.3.5. Klasik Test Teorisinin Dezavantajları**

KTT'nin dezavantajları aşağıdaki gibidir:

1. KTT'de madde ve ölçme aracı istatistikleri elde edildikleri örnekleme bağımlıdır. Örneğin, ölçme aracının güvenilirliğinin hesaplanmasında, ölçme aracının tekrar uygulanması gereklidir. Bireylerin  $\theta$  düzeyleri, iki uygulama arasında yeni öğrenmeler, unutmalar, kaygı, motivasyon, hatırlama gibi bazı etkilerle değiştiği için bu istatistiklerin genellenmesi mümkün olmamaktadır.

2. KTT'de bireylerin  $\theta$  düzeyleri; uygulanan ölçme aracına, dolayısıyla ölçme aracının maddelerine bağımlıdır. Bir birey, aynı özelliği ölçen farklı zorluktaki ölçme araçlarından farklı puanlar alır. Bu nedenle bireylerin karşılaştırılması sadece aynı ya da paralel ölçme araçları kullanıldığında mümkün olabilmektedir.
3. KTT ile geliştirilen ölçme araçları, genellikle  $\theta$ 'nın orta düzeyindeki bireylere hitap eder. Bu durumda, alt ve üst düzeydeki bireylerin  $\theta$  düzeyleri güvenilir şekilde belirlenemez.
4. KTT'de, ölçme hatasına ilişkin varyansın tüm bireyler için eşit olduğu varsayılır.
5. Yanıt kategorileri arasındaki uzaklıkların aynı olduğu varsayılır.

KTT'nin sınırlılıklarını aşma çalışmaları, MYT'nin gelişmesiyle sonuçlanmıştır (Anıl, 2002; Doğan, 2002).

#### **1.4. Madde Yanıt Teorisi (Item Response Theory)**

##### **1.4.1. Madde Yanıt Teorisinin Temel Özellikleri**

MYT, 1930'lu yılların sonunda KTT'nin sınırlılıklarını giderebileceği iddiası ile ortaya atılmıştır. Bu teori, "örtük özellikler teorisi", "madde tepki teorisi" ya da "modern test teorisi" adlarıyla da anılmaktadır. Teori, 1950–1970 yılları arasında matematiksel işlemlerinin karmaşıklığı nedeniyle yavaş gelişmiş, bilgisayar teknolojisindeki gelişmelere paralel olarak teorinin kullanımı ve gelişimi de hızlanmıştır.

MYT, örneklemden bağımsız madde kalibrasyonunu ve maddeden bağımsız olarak bireyin  $\theta$  düzeyini tahmin etmeyi olanaklı hale getirmek iddiasındadır (Doğan, 2002). Hambleton ve ark.'na (1991) göre MYT iki temel gerçeğe dayanır: a) Bireyin ölçme aracında yer alan bir maddedeki

performansı özellik ya da gizli özellik olarak adlandırılan bir grup faktör ile açıklanabilir; b) Bireyin maddedeki performansı ve madde performansının altında yatan özellikler seti arasındaki ilişki MKF olarak adlandırılan tekdüze artan bir fonksiyon ile tanımlanabilir. Bu fonksiyon, MYT'de seçilecek olan modele bağlı olarak uygun parametreler ile ifade edilir. Bu parametreler; bir parametrelili modelde madde zorluk parametresi, iki parametrelili modelde madde zorluk parametresi ve madde ayırt edicilik parametresi, üç parametrelili modelde ise madde zorluk parametresi, madde ayırtedicilik parametresi ve şansa bağlı tahmin parametresidir.

KTT, ölçme aracını oluşturan maddelerden elde edilen toplam puana dayalı iken; MYT ölçme aracını oluşturan maddelere dayalıdır. KTT'de asıl inceleme konusu bireyin ham puanı iken; MYT'de bireyin ölçme aracındaki her bir maddeye verdiği yanıtıdır (Embretson ve Reise, 2000; Baker, 2001).

#### **1.4.2. Madde Yanıt Teorisinin Varsayımları**

MYT modellerinin temel varsayımları aşağıdaki gibidir:

1. Tek Boyutluluk (Unidimensionality): Tek boyutluluk, ölçme aracının tek bir boyutu, tek bir değişkeni ölçmesi anlamındadır. Tek boyutluluk pratikte sağlanması zor bir varsayım olup; bu boyutun ne olduğunu tahmin etmek ve boyutu tanımlamak her zaman kolay olmayabilir. Tek boyutluluğun incelenmesinde faktör analizinden, ilişki katsayılarından ve yamaç eğim grafiğinden (scree plot) faydalanılır (Hambleton ve Swaminathan, 1985).
2. Yerel Bağımsızlık (Local Independence): Yerel bağımsızlık, bireylerin  $\theta$  düzeyleri sabit tutulduğunda, ölçme aracının maddelerine verdikleri yanıtların istatistiksel olarak bağımsız olmasıdır. Lord (1980), tek boyutluluk varsayımının sağlanması durumunda belli bir  $\theta$  düzeyine sahip olan bireyin

maddelere vermiş olduğu yanıtlar arasındaki ilişki katsayısının sıfır olduğunu ve dolayısıyla tek boyutlu olan ölçme araçlarının aynı zamanda yerel bağımsızlık varsayımını da sağladığını ileri sürmüştür. Bu düşünceden yola çıkarak, Hambleton ve Swaminathan (1985), yerel bağımsızlığın test edilmesi için de faktör analizinin kullanılabileceğini belirtmiştir. Bunun yanı sıra madde puanları arasındaki tetrakorik ilişki katsayıları veya ki-kare analizi kullanılabilecek diğer yöntemlerdir.

3. Normal Dağılım (Normal Distribution): MYT modellerinin tümü için ham puanların normal dağılıma sahip olması gereklidir. Verilerin normal dağılıma uygun olup olmadığının incelenmesinde normallik testlerinden ve grafiksel yöntemlerden yararlanır (Öztuna ve ark., 2006). Ancak yapılan araştırmaların birçoğunda, normal dağılım varsayımı sağlanmasa bile, yukarıdaki iki varsayımın sağlanması durumunda çalışmalara devam edildiği görülmüştür (Doğan, 2002).

4. Ölçme Aracının Uygulaması İçin Verilen Sürenin Yeterli Olması (Non-Speeded Test Administration): Ölçme aracının “speeded” olması, bazı bireylerin tüm maddeleri yanıtlamak için yeterli zamana sahip olmaması anlamına gelir (Goegebeur ve ark., 2006). MYT’de ise ölçme aracının “speeded” olmaması gerekir, yani bireylerin hepsi tüm maddeleri yanıtlamak için yeterli zamana sahip olmalıdır. Bu varsayım, yanıtlanmayan madde sayısının varyansının, yanlış yanıtlanan madde sayısının varyansına oranı ile test edilir ve bu oran sıfıra yakınsa varsayımın sağlandığı söylenir (Hambleton ve ark., 1991).

### 1.4.3. Madde Yanıt Teorisinde Madde ve Ölçme Aracı İstatistikleri

#### 1.4.3.1. Madde Parametreleri

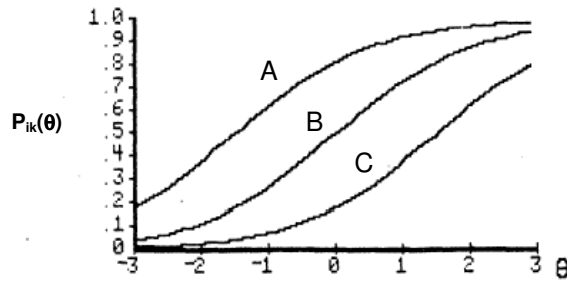
MYT'de  $\theta$ , madde yanıtları arasındaki kovaryansı açıklayan sürekli ve tek boyutlu bir yapı olarak tanımlanır (Reeve, 2002).  $\theta$ , eğitim alanında "yetenek/bilgi" olarak; sağlık alanında ise kullanılan ölçme aracına bağlı olarak "fonksiyonellik, özürlülük, depresyon ya da yaşam kalitesi" olarak adlandırılabilir.  $\theta$ , lojit birimi ile ifade edilir ve kuramsal olarak  $-\infty$  ile  $+\infty$  arasında değer alır.

Bireye ilişkin  $\theta$  düzeyi ile  $i$ . maddenin  $k$ . kategorisini seçme olasılığı  $[P_{ik}(\theta)]$  arasındaki ilişki MKF ile ifade edilir. MKF, örtük özellikler uzayı tek boyutlu ise monoton artan bir fonksiyon olup;  $\theta$  arttıkça,  $P_{ik}(\theta)$ 'nin artacağını gösterir ve bu varsayım MYT'nin temelini oluşturur. MYT'de bir maddenin özelliklerini incelemeye kullanılan ve MKF'nin şeklini belirleyen üç temel parametre vardır:

1. Madde zorluk parametresi (item difficulty parameter- $b$ ): Bir maddenin  $b$  değeri, maddenin  $\theta$  düzeyi boyunca hangi nokta/aralıkta iyi işlediğini gösterir. Örneğin; kolay bir madde  $\theta$  düzeyi düşük olan bireyler için; zor bir madde ise  $\theta$  düzeyi yüksek olan bireyler için daha iyi işler.  $b$  değeri,  $\theta$  ile ilişkili olarak MKF eğrisinin konumunu belirlediği için "konum parametresi" olarak da adlandırılır. Parametre değerinin artması MKF eğrisinin sağa doğru, azalması ise sola doğru kaymasına neden olur.  $b$  değeri, kuramsal olarak  $-\infty$  ile  $+\infty$  aralığında değer alabilir (Hambleton ve Swaminathan, 1985; Baker, 2001).  $b$  değerinin negatif olması maddenin kolay, pozitif olması maddenin zor, 0'a yakın olması ise maddenin orta güçlükte olduğuna işaret eder.

2. Madde ayırtedicilik parametresi (item discrimination parameter-a): Bu parametre bir maddenin,  $\theta$  düzeyi b değerinin altında olan bireylerle, üstünde olan bireyleri ne derece iyi ayırdığını gösterir. a değeri, MKF eğrisinin eğimine (dikliğine) karşılık gelmektedir. Eğri dikleştikçe maddenin ayırtediciliğinin arttığı, eğri düzleştikçe, maddenin ayırtediciliğinin düştüğü söylenir. Parametre değeri yükseldikçe ayırtedicilik arttığı için maddenin; ölçme aracının bütünü ile ölçülen özelliğe sağladığı bilgi miktarı da artmaktadır. Bu bakımdan a, “madde geçerlik ölçüsü” olarak da tanımlanabilir. a değeri, kuramsal olarak  $-\infty$  ile  $+\infty$  aralığında değer alabilir. Bazı çalışmalar a değeri 0.80’in altında olan (Baker, 2001), bazıları ise 0’ın altında olan (Wiberg, 2004) maddelerin kullanılmamasını önermektedir. a değeri negatif olan maddelerde ise, doğru yanıt verme olasılığı  $\theta$ ’nın alt düzeylerinde üst düzeylerine göre daha yüksektir. Bu tür maddelerin ölçülen özellik dışında bir özelliği ölçtüğü veya hatalı olduğu kabul edilip; ölçme aracından çıkarılması gerekmektedir.

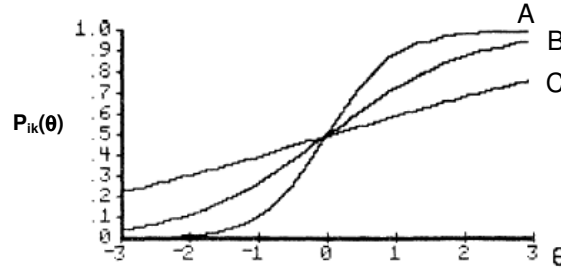
Şekil 1.1’de, üç farklı madde için MKF eğrileri gösterilmektedir. Şekildeki tüm maddeler için a değerleri aynı olmasına rağmen, b değerleri değişkenlik göstermektedir. En soldaki eğri (A) “kolay” bir maddeyi; ortadaki eğri (B) “orta zorlukta” bir maddeyi ve en sağdaki eğri de (C) “zor” bir maddeyi temsil etmektedir.



Şekil 1.1: Ayırt edicilik parametreleri aynı, zorluk parametreleri farklı olan üç maddeye ilişkin MKF eğrileri



Şekil 1.2'de, b değerleri aynı olmasına rağmen, a değerleri farklı olan üç maddeye ait MKF eğrileri gösterilmektedir. En üstteki eğri (A), yüksek ayırt ediciliğe sahip bir maddeye, ortadaki eğri (B), orta düzeyde ayırt ediciliğe sahip bir maddeye, üçüncü eğri ise (C) düşük ayırt ediciliğe sahip olan bir maddeye aittir.



Şekil 1.2: Zorluk parametreleri aynı, ayırt edicilik parametreleri farklı olan üç maddeye ilişkin MKF eğrileri

3. Şansa bağlı tahmin parametresi (guessing parameter-c): Bir maddenin doğru yanıtlanması için gereken en düşük  $\theta$  değeridir. Doğru yanıtı bilmeyen birey, c olasılığı ile doğru yanıtı tahmin edecektir (Baker, 2001). c değeri, 0 ile 1 aralığında değer alabilir (Anıl, 2002).

#### 1.4.3.2. Madde ve Ölçme Aracı Bilgisi (Item and Test Information)

KTT'deki güvenilirlik kavramına karşılık, MYT'de madde ve ölçme aracı bilgi fonksiyonları hesaplanır. Bir madde tarafından sağlanan bilgi, o maddenin  $\theta$  düzeyi boyunca verilen herhangi bir noktada/aralıkta,  $\theta$ 'nın tahmin edilmesine sağladığı katkı olarak ifade edilir.  $I_i(\theta)$ ,  $\theta$ 'nın bir fonksiyonu olduğundan, seçilen her  $\theta$  düzeyinde farklı değerler alır. Böylece ölçmenin duyarlılığı  $\theta$ 'nın her düzeyi için saptanabilir. Ayrıca bu fonksiyon sayesinde her maddenin ölçme aracının duyarlılığına ne derece katkıda bulunduğu saptanıp, bir ölçme aracı oluştururken madde seçimi buna göre yapılabilir.

Ölçme aracı bilgi fonksiyonu  $[I(\theta)]$  ise,  $I_i(\theta)$  değerlerinin toplamıdır. Ölçme aracı tarafından sağlanan bilgi arttıkça,  $\theta$  düzeyine ilişkin tahminin standart hatası azalır. Tahminin standart hatası  $[SE(\theta)]$  aşağıdaki gibi hesaplanır:

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}} \quad (\text{Denklem 1.1})$$

$SE(\theta)$ 'nin büyüklüğü, genel olarak (i) ölçme aracında yer alan madde sayısına (fazla sayıda maddeden oluşan ölçme araçları, daha düşük  $SE(\theta)$  değerine sahiptir), (ii) b değeri ile  $\theta$  arasındaki uyuma (oldukça zor ya da kolay maddelerden oluşan ölçme araçlarının tersine, b değeri bireyin  $\theta$  düzeyine eşit ya da benzer olan maddelerden oluşan bir ölçme aracı, daha düşük  $SE(\theta)$  değerine sahiptir) bağlıdır (Hambleton ve ark., 1991).

#### 1.4.4. Madde Yanıt Teorisinde Kullanılan Modeller

MYT modelleri genel olarak puanlama formatına (yanıt kategorilerine) göre sınıflandırılır. Ölçme sonucunda maddelerden elde edilebilecek iki tip format vardır:

- İki sonuçlu (dichotomous) puanlama
- Çok sonuçlu (polytomous) puanlama

MYT, gelişmeye başladığı 1950'li yıllarda daha çok iki sonuçlu puanlamaya sahip ölçme araçlarının geliştirilmesi ve değerlendirilmesinde kullanılmıştır. Uygulamalarda, bireylerin yanıtlarının ikiden fazla kategoride incelenmesi durumunun gündeme gelmesi ile başka modellere de gereksinim duyulmuştur. Çok sonuçlu puanlama modellerinde en iyi yanıt dışındaki diğer

kategori seçimlerine kısmi puan verilirken; iki sonuçlu modellerde hiç puan verilmez. Puanlama formatına göre MYT modelleri, Çizelge 1.2'de sunulmuştur.

Çizelge 1.2. Puanlama Formatına Göre MYT Modelleri

<b>Puanlama Formatı</b>	<b>MYT Modeli</b>
İki Sonuçlu	Bir Parametrel Model
İki Sonuçlu	Rasch Modeli*
İki Sonuçlu	İki Parametrel Model
İki Sonuçlu	Üç Parametrel Model
İki Sonuçlu	Dört Parametrel Model
Çok Sonuçlu	Derecelendirilmiş Yanıt Modeli (Graded Response Model)
Çok Sonuçlu	İsimsel Model (Nominal Model)
Çok Sonuçlu	Kısmi Kredi Modeli (Partial Credit Model)*
Çok Sonuçlu	Sıralı Sonuçlu Model (Rating Scale Model)*
Çok Sonuçlu	Genelleştirilmiş Kısmi Kredi Modeli (Generalized Partial Credit Model)

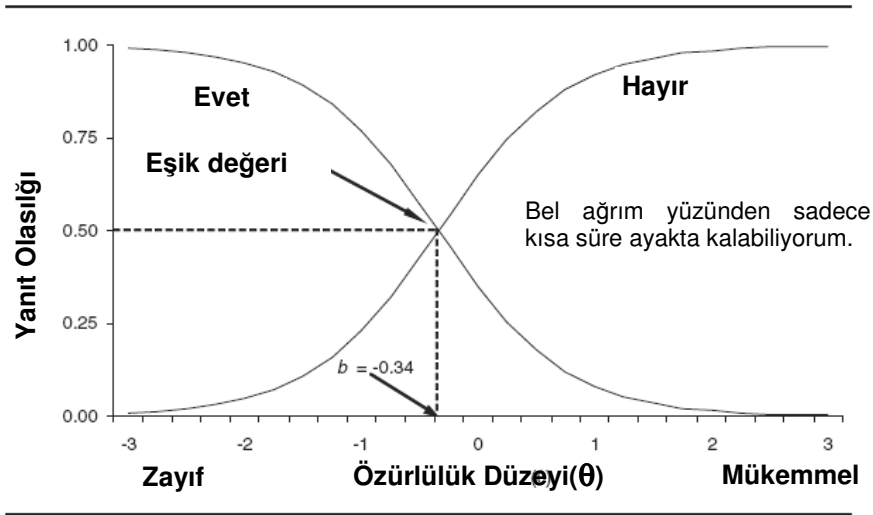
\* Rasch modelleri ailesindedir.

Thissen ve Orlando (2001), MYT'de model yapılandırma amacıyla kullanılan iki yaklaşımı tartışmışlardır. Bunlardan ilki, veriye iyi uyum sağlayacak modelleri geliştirmektir. Bu yaklaşımın amacı, madde analizidir. Model, verinin özelliklerini yeterli ve kesin bir şekilde yansıtmalıdır. Madde yanıt kategorisi iki sonuçlu olduğunda bir, iki veya üç parametrel modeller kullanılırken; çok sonuçlu olduğunda derecelendirilmiş yanıt modeli veya genelleştirilmiş kısmi kredi modeli kullanılmakta, bahsedilen modeller ilk yaklaşımı temel almaktadır. Model yapılandırma yaklaşımı, ölçüm teorisinin veriyi açıklamak olduğu düşüncesine dayalıdır. İkinci yaklaşım, verinin özel ölçüm özellikleri tanımlanan modele uyum göstermesini sağlamaktır. Madde ya da birey, modele uyum sağlamazsa, uyumsuz madde ya da birey analizden çıkartılır. Madde yanıt kategorisi iki sonuçlu olduğunda ikili Rasch modeli kullanılırken; çok sonuçlu olduğunda kısmi kredi modeli veya sıralama ölçekli model kullanılmakta, bahsedilen modeller ikinci yaklaşımı temel almaktadır.

MKF'ye göre tanımlanan, "normal ogive" ve "lojistik" olarak adlandırılan iki model grubu vardır. İki model arasında madde parametrelerinin yorumlanması bakımından farklılık olmazken; uygulamada matematiksel kolaylıklar sağlaması bakımından lojistik model daha çok tercih edilmektedir (Wright ve Stone, 1979).

#### 1.4.4.1. İki Sonuçlu Puanlama Modelleri

İki sonuçlu MYT modelleri arasındaki temel ayrım, maddeleri tanımlamak için kullanılan parametre sayısıdır. Şekil 1.3, iki sonuçlu puanlamaya sahip bir maddenin MKF'sini göstermektedir (Chang ve Reeve, 2005).



Şekil 1.3. İki Sonuçlu Puanlamaya Sahip Bir Madde için MKF

**Bir Parametrelili Model:** Bu modelde a değeri, tüm maddeler üzerinden a değerlerinin ortalamasına eşit olup; b değeri her bir madde için değişkenlik gösterir.

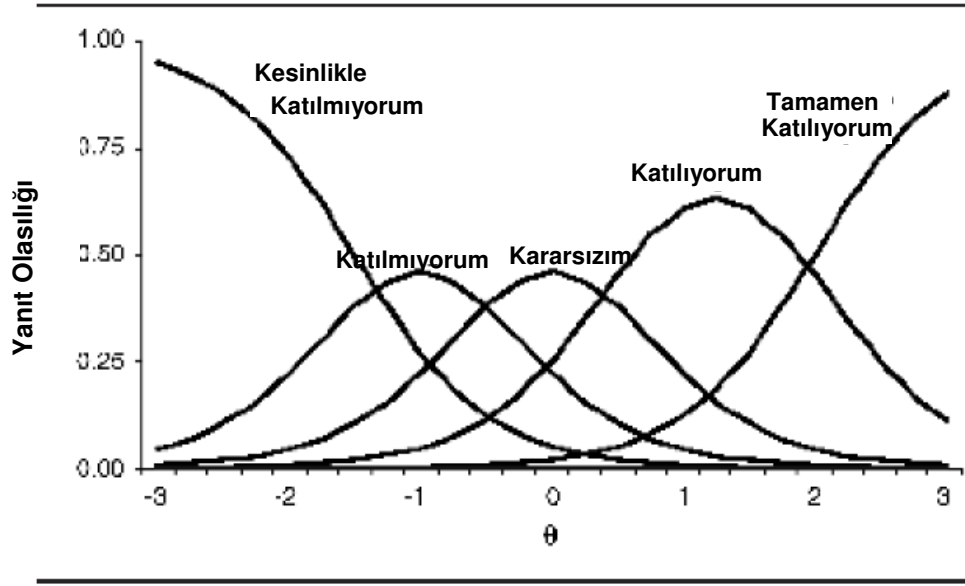
**Rasch Modeli:** Danimarkalı matematikçi Georg Rasch tarafından 1960 yılında ortaya koyulan bu model, matematiksel olarak bir parametrelili modelle benzer bir yapıya sahiptir. Farklılık  $a$  değerinin 1 olarak alınmasıdır. Rasch modelinin önemli bir özelliği, gözlenen ham puanların, doğrusal olmayan bir dönüşüm kullanarak  $\theta$  puanlarının tahmin edilmesi için yeterli bir istatistik olmasıdır. MYT'nin diğer modellerinde, ham puan yeterli istatistik değildir (Hays ve ark., 2000).

**İki Parametrelili Model:** İki parametrelili model,  $\theta$ 'nın tahmin edilmesinde tek parametrenin yetersiz olduğu görüşünden doğmuştur. Bu modelde  $b$  parametresine ek olarak,  $a$  parametresi de bulunmaktadır. İki parametrelili modelin önemli bir özelliği, bireyin  $\theta$  düzeyi ile  $b$  değeri arasındaki farklılığın,  $a$  değeri yüksek olan maddelerin yanıtlanması olasılığı üzerinde önemli bir etkiye sahip olmasıdır.

**Üç ve Dört Parametrelili Model:** Üç parametrelili modelde,  $a$  ve  $b$  parametrelerine ek olarak,  $c$  parametresi yer alırken; dört parametrelili modeldeki üç parametreye ek olarak  $\theta$  düzeyi yüksek olan bireylerin, çok kolay bir maddeyi yanlış yanıtlama olasılığını hesaba katan bir parametre de yer almaktadır.

#### **1.4.4.2. Çok Sonuçlu Puanlama Modelleri**

Çok sonuçlu puanlanan maddeler, bireyin  $\theta$  düzeyi hakkında daha fazla bilgi sağladığından oldukça avantajlıdır. Çok sonuçlu puanlama modelleri kullanılarak bireyin  $\theta$  düzeyi ile belirli bir yanıt kategorisini seçme olasılığı arasındaki doğrusal olmayan ilişki açıklanmaya çalışılır. Şekil 1.4, çok sonuçlu puanlamaya sahip örnek bir maddenin MKF'sini göstermektedir (Chang ve Reeve, 2005).



Şekil 1.4. Çok Sonuçlu Puanlamaya Sahip Bir Madde için MKF

**Derecelendirilmiş Yanıt Modeli:** İki parametrelili lojistik modelin, çok sonuçlu puanlama durumunda kullanılmak üzere geliştirilmiş halidir. Samejima tarafından 1969 yılında önerilen modelde, yanıt kategorileri Likert türü ölçeklerde olduğu gibi belirli bir sıraya sahiptir. Derecelendirilmiş yanıt modelinde, her madde tek bir “a değeri” ve “eşik değerleri” ile tanımlanır. Derecelendirilmiş yanıt modelinde, her bir eşik değeri için bir tane MKF tahmin edilir. (Hays ve ark., 2000).

**İsimsel Model:** Çok sonuçlu puanlama durumu için 1972 yılında Bock tarafından önerilen bu modelde, yanıt kategorileri arasında belirli bir sıralama yoktur. Bu modelin amacı, maddelerin doğru yanıtlanıp yanıtlanmadığının yanı sıra, bireylerin yanıtlarından elde edilen bilginin tümünü kullanarak,  $\theta$ 'ya ilişkin tahmin değerinin kesinliğini artırmaya çalışmaktır. Her bir yanıt kategorisi için ayrı bir b ve a değeri hesaplanır (Hambleton ve ark., 1991).

**Kısmi Kredi Modeli:** Bu model, çoklu aşama gerektiren ve çözümlenme sürecinde farklı aşamaların tamamlanması durumunda kısmi kredi vermenin önemli olduğu veya Likert tipi maddelerde yanıt kategorileri arasındaki uzaklıkların farklılık gösterdiği ölçme aracı maddelerinin analiz edilmesi için Masters tarafından 1982 yılında geliştirilmiştir (Chang ve Reeve, 2005). Rasch modelinin çok sonuçlu puanlama için genelleştirilmiş biçimi olması nedeniyle,  $a$  değeri tüm maddeler için eşittir. Model, bireyin  $x$  kategorisinde yanıt verme olasılığını,  $\theta$  ve kategori eşik değeri arasındaki farklılığın bir fonksiyonu olarak tanımlar. Kısmi kredi modeli, yanıt kategorilerinin sırası hakkında hiçbir varsayım yapmaz.

**Sıralı Sonuçlu Model:** Andrich tarafından 1978 yılında geliştirilen bu model, iki sonuçlu Rasch modelinin bir uzantısıdır (Chang ve Reeve, 2005). Bu modelde yanıt kategorileri belirli bir sıraya sahiptir. Yanıt kategorileri, her bir madde için eşit olarak değerlendirilen eşik değerlerine ayrılır ve  $b$  değeri tek bir parametre ile tanımlanır.  $b$  değeri, seçilen bir madde için kategori eşik değerlerine göre ortalama zorluğu gösterir. Tüm maddeler için  $a$  değeri eşittir (Hays ve ark., 2000).

Sıralı sonuçlu model ve kısmi kredi modeli arasındaki en temel farklılık, sıralı sonuçlu modelde aynı yanıt formatının ölçme aracındaki tüm maddeler için geçerli olduğunun ve eşik değerleri arasındaki farklılığın tüm maddeler için eşit olduğunun varsayılmasıdır. Örneğin, birinci ve ikinci kategoriyi birbirinden ayıran eşik değeri ile ikinci ve üçüncü kategoriyi birbirinden ayıran eşik değeri arasındaki uzaklık, tüm maddeler için eşittir.

**Genelleştirilmiş Kısmi Kredi Modeli:** Muraki tarafından 1992 yılında geliştirilen model, her bir çok sonuçlu kategoriye sahip maddeyi, iki parametre cinsinden tanımlar: maddenin her bir kategorisi için  $a$  ve  $b$  değeri. Bu model, kısmi kredi modelinin genelleştirilmiş şekli olduğu için,  $a$  değerleri maddelere göre farklılık gösterir.

### 1.4.5. İncelenen Özelliğin ve Madde Parametrelerinin Tahmini

$\theta$  ve madde parametrelerinin tahmini, MYT'nin uygulanmasında en önemli adımdır. MYT modellerinde  $P_{ik}(\theta)$ , bireyin  $\theta$  düzeyine ve madde parametrelerine bağlıdır. Bu parametrelerin hiçbiri başlangıçta bilinmemekle beraber, birbirleriyle ilişkilidir. Bu yüzden, tüm parametreler eş zamanlı olarak tahmin edilir. Bu amaçla, bireylerin ölçme aracında yer alan maddelere verdikleri yanıtlar üzerinden olabirlik fonksiyonu (L) hesaplanır ve bu fonksiyonu maksimum yapan  $\theta$  ve madde parametreleri tahmin edilir.

En Çok Olabirlik (Maximum Likelihood-ML) yöntemi,  $\theta$  ve madde parametrelerinin tahmini için en yaygın kullanılan yöntemdir. En çok olabirlik tahminlerini hesaplamak, farklı en az iki kategoride yanıt verildiğinde mümkündür. Bu aşamada en çok olabirlik tahmin edicileri, L'nin  $\theta$  ve madde parametrelerine göre ayrı ayrı türevlerinin alınıp sıfıra eşitlenmesinden elde edilir.

Birleşik en çok olabirlik tahmin yönteminde,  $\theta$  ve madde parametreleri, eş zamanlı olarak tahmin edilir. Marjinal en çok olabirlik yönteminde bireylerin  $\theta$  düzeylerinin önsel bir dağılıma sahip olduğu varsayılır ve madde parametreleri bu dağılıma dayalı olarak tahmin edilir. Koşullu en çok olabirlik yöntemi, Rasch tipi modellere uygulanabilir. Bu yöntemde olabirlik fonksiyonu, Rasch modelinde yeterli bir istatistik olan toplam doğru puan üzerine koşullandırılır ve model "sadece parametreyi içeren" ve "sadece yeterli istatistiği içeren" bölümlere ayrılır.

En çok olabirlik tahmin edicileri asimtotik yansızlık, asimtotik etkinlik ve asimtotik normallik gibi bir çok kriter bakımından en popüler tahmin edicilerdir. Eğer L'nin ilgili parametrelere göre türevlerinden elde edilen denklemler, doğrusal fonksiyonlardan oluşuyorsa açık olarak ifade edilebilen



tek bir çözüme sahiptirler. Bununla beraber, denklemler doğrusal olmayan fonksiyonlardan oluşuyorsa bu denklemlerin matematiksel çözümleri bulunamaz. Dolayısıyla bu denklemleri çözmek için iteratif yöntemler kullanmak zorunlu hale gelir. Bu amaçla kullanılan en yaygın yöntem Newton-Raphson algoritmasıdır (Mathews ve Fink, 2004). Bununla birlikte, iteratif yöntemler kullanarak çözüme ulaşmak her zaman mümkün olmayabilir. Çünkü birden fazla kök, iterasyonların yakınsamaması veya yanlış değerlere yakınsama gibi problemlerle karşılaşma ihtimali yüksektir (Barnett, 1966; Tiku, 1967; Vaughan, 1992).

Bireylerin  $\theta$  düzeylerinin ve madde parametrelerinin tahmin edilmesinde kullanılmak üzere geliştirilmiş Bayesçi yöntemler de mevcuttur. Bu yöntemlerde  $\theta$  ve madde parametrelerine ilişkin önsel dağılımlar belirlenir. Bu önsel bilgiler kullanılarak sonsal  $\theta$  ve madde parametreleri tahmin edilir (Hambleton ve ark., 1991). Bayesçi yöntemler,  $\theta$  düzeyine ilişkin bir tahmin yapılmadan önce en azından iki farklı kategoride yanıt gereksinim duyan en çok olabilirlik yönteminin tersine, çok az sayıda maddeden sonra tahmin elde edilebilme avantajına sahiptir. Bununla beraber, Bayesçi tahminlerin önemli bir dezavantajı; tahminlerin seçilen önsel olasılıktan kaynaklanan bir yanlılığa sahip olmasıdır.

#### **1.4.6. Model-Veri Uyumunun ve Madde İşlev Farklılığının Belirlenmesi**

##### **1.4.6.1. Model-Veri Uyumunun Belirlenmesi**

Literatürde yapılan çalışmalarda, model-veri uyumu ve varsa uyumsuzluğun nedenleri yeterince incelenmemiş; incelemenin yapıldığı çalışmalarda ise uygun olmayan istatistiklerin kullanıldığı görülmüştür. Bu çalışmalarla ilgili diğer bir problem, model uyumu için kullanılan istatistiksel testlere fazlaca

güvenilmesidir. Gerçekte bu testler, birey sayısına çok duyarlıdır. Örneklem büyüklüğü yeterli olduğunda, incelenen modelden ayrılıkların hemen hepsi, model-veri uyumu ile ilgili yokluk hipotezinin reddedilmesi ile sonuçlanacaktır. Örneklem büyüklüğü çok küçük olduğunda ise, büyük model-veri ayrılıkları, testin gücünün düşük olması nedeniyle belirlenemeyecektir. Bunun yanında, örneklem büyüklüğüne dayalı olan parametre tahminleri, yüksek  $SE(\theta)$  değerine sahip olduğu için kullanılamayacaktır. Ayrıca, bazı uyum iyiliği istatistiklerinin örneklem dağılımları iddia edilenden farklı olduğunda, bu istatistikler bilinen dağılımların belirlenen değerleri kullanılarak yorumlandığında hatalı sonuçlar verebilecektir.

Hambleton ve Swaminathan (1985), modelin veriye uyumu hakkında çıkarsamada bulunmak için 3 farklı inceleme yapılmasını önermiştir:

1. Ölçme aracı verisi için model varsayımlarının geçerliğinin incelenmesi:  
Bölüm 1.4.2'de bahsedilen varsayımların sağlanıp sağlanmadığı incelenir.

2. Modelin beklenen özelliklerinin ( $\theta$  ve madde parametre tahminlerinin değişmezliği) incelenmesi: Farklı maddelerden (örneğin, zor ve kolay) oluşan ölçme araçları için  $\theta$  tahminlerinin, farklı gruplar (örneğin, erkek ve kız) için madde parametrelerinin değişkenlik göstermemesi istenir.

3. Gerçek veriyi, uygunsa ölçme aracından elde edilen veriyi kullanarak model tahminlerinin doğruluğunun incelenmesi:

*Artık değerlerinin incelenmesi:* Bu inceleme için, seçilen MYT modeli kullanılarak  $\theta$  düzeyi ile madde parametreleri tahmin edilir ve  $\theta$  düzeyi, eşit genişlikteki aralıklara bölünür. Gözlenen doğru yanıt oranı,  $\theta$ 'nın her bir kategorisinde maddeyi doğru yanıtlayan birey sayısının o kategorideki birey sayısına bölünmesi ile hesaplanırken; beklenen doğru yanıt oranı incelenen model kullanılarak elde edilir.

Seçilen modelin geçerli olduğu varsayımı altında, gözlenen doğru yanıt oranı ile beklenen doğru yanıt oranı arasındaki fark, yani “ham artık”,  $\theta$ 'nın alt grupları için hesaplanır. Ham artığın bir sınırlılığı, her bir kategorideki beklenen oran ile ilişkili örnekleme hatasını dikkate almamasıdır. Örnekleme hatasını hesaba katmak için ham artık, beklenen doğru oranın standart hatasına bölünerek “standartlaştırılmış artık” değeri hesaplanır. Artık değerleri küçük ve MKF eğrisi boyunca rasgele bir dağılım gösteriyorsa, eğrinin veriye uyduğu sonucuna ulaşılabılır. MYT modelleri seçilirken; farklı modeller için artıkların, standartlaştırılmış artıkların veya her ikisinin de incelenmesi önemli bilgi verebilir.

Model-veri uyumunu değerlendirmenin diğer bir yolu, ki-kare test istatistiğinin incelenmesidir. İstatistiğin gözlenen değeri, kritik değeri aşarsa, eğrinin veriye uyduğu biçimindeki yokluk hipotezi reddedilir ve daha iyi uyum gösteren bir model bulunmaya çalışılır (Hambleton ve ark., 1991).

Modelin beklentilerine uymayan maddeler, ölçme aracı ile uyum sağlamadığı düşünülerek ölçme aracından çıkarılır veya üzerinde düzeltme yapılmaya çalışılır. Uyum analizleri sonucunda çıkarılan bir maddenin; yazımında teknik hata olduğu,  $\theta$ 'dan farklı bir değişkeni ya da boyutu ölçtüğü, ölçme aracındaki diğer maddelerden daha zor olduğu veya  $\theta$  dağılımı çok geniş bir kitleye uygulandığı düşünülebilir.

#### **1.4.6.2. Madde İşlev Farklılığının Belirlenmesi**

Madde işlev farklılığı (MİF), bir maddenin bir grupta diğer gruba göre farklı işlemesidir. Diğer bir ifadeyle,  $\theta$  boyunca belirli düzeylere sahip olan; fakat farklı kitlelerden gelen bireylerin bir maddeye yanıt verme olasılıklarının farklı olmasıdır. İşlev farklılığı gösteren maddeler içeren ölçme araçları, gruplararası karşılaştırma için düşük geçerliğe sahip olabilirler, çünkü araçtan

elde edilen puanlar ölçme aracının ölçmesi istenenden başka özellik türlerinin göstergesi olabilir (Thissen ve ark., 1988).

MIF, iki ya da daha fazla grubun MKF'lerinin farklı yollarla karşılaştırılması ile incelenebilir. İlki ve en yaygın kullanılanı, MKF eğrilerini tanımlayan parametrelerin karşılaştırılmasıdır. MKF eğrileri, madde parametreleri ile belirlendiği için iki eğri sadece madde parametreleri bakımından farklılık gösterecektir. İkinci yaklaşım, MKF eğrileri arasındaki alanın değerlendirilmesidir. MKF eğrileri arasındaki alan sıfıra eşit ise, eğriler çakıştığı için MIF olmadığı sonucuna ulaşılabacaktır (Hambleton ve ark., 1991).

#### **1.4.7. Madde Yanıt Teorisinin Avantajları**

MYT'nin önemli bir avantajı, değişmezlik özelliğidir (invariance). Değişmezlik, ölçme aracı ve madde parametrelerinin uygulandığı gruptan bağımsız olması, bireylerin  $\theta$  düzeylerinin ise uygulanan ölçme aracından bağımsız olmasıdır (Hambleton ve Swaminathan, 1985). Ayrıca,  $I_i(\theta)$  ve  $I(\theta)$  değerlerinin de değişmediği iddia edilmektedir. Böylece maddelerin ve ölçme aracının uygulandığı çeşitli grupların karşılaştırılması mümkün olmaktadır. Değişmezlik özelliği,  $\theta$ 'nın tahmin edilmesi sürecinde madde parametreleri bilgisinin, madde parametrelerinin tahmin sürecinde ise  $\theta$  bilgisinin dahil edilmesi ile elde edilir.

MYT'nin, bahsedilen avantajların dışında, ölçme aracı geliştirmede, soru bankası ve uyarlamalı ölçme aracı oluşturmada, ölçme aracı maddelerinin belirli bir grup için yanlılık gösterip göstermediğini belirlemede, madde ve seçenekleri ağırlıklandırmada ve ölçme araçlarının eşitlenmesinde karşılaşılan sorunlara çözüm getirdiği iddia edilmektedir. Özellikle bilgisayarda ölçme uygulamaları ile bireysel ölçme araçları için en uygun teori, MYT olarak görülmektedir.

#### 1.4.8. Madde Yanıt Teorisinin Dezavantajları

MYT'nin en önemli sınırlılığı, teorisin varsayımlarını karşılamamanın gerçek verilerde zor olmasıdır. Diğer bir dezavantaj, MKF'nin tanımlanmasındaki karmaşıklık nedeniyle parametrelerin hesaplanmasının zor olmasıdır. Bu nedenle, özellikle madde ve birey sayısı fazla olduğunda, bilgisayar programına ihtiyaç duyulmaktadır. Bu amaçla hazırlanmış çok sayıda özel yazılım bulunmaktadır (BILOG, MULTILOG, PARSCALE, RUMM gibi).

MYT modellerinin uygulanmasında büyük örneklemelere gereksinim duyulması, teorisin yaygın kullanımını sınırlayan bir faktör olabilir. Örneklem büyüklüğü, kullanılan ölçüm modeline (Rasch türü modeller, daha az sayıda parametre tahmin ettiği için, diğer MYT modellerine göre daha küçük örneklem büyüklüklerine gereksinim vardır), çalışmanın amacına (Amaç, bireylerin  $\theta$  düzeylerinin tahmininden daha çok, uyum istatistiklerinin hesaplanması ise, daha fazla sayıda bireye ihtiyaç vardır (Downing, 2003)), bireylerin  $\theta$  düzeylerinin örnekleme dağılımına ve yanıt kategorilerinin sayısına (Yanıt kategorilerinin sayısı arttıkça, tahmin edilecek parametre sayısı da artacağından büyük örneklemelere olan ihtiyaç da artar) bağlıdır.

#### 1.4.9. Ölçme Aracı Geliştirme Açısından İki Teorinin Karşılaştırılması

MYT, "modern test teorisi" olarak adlandırılırsa da, psikologlar ve eğitimciler tarafından kullanılan ölçme araçlarının çoğu KTT ile geliştirilmiştir. İki yaklaşım arasında bireylerin yanıt profillerinin işleyişi kadar, değerlendirme ve ölçme aracı geliştirme bakımlarından da önemli farklar vardır. MYT ile KTT'nin karşılaştırılması sonucunda aşağıdaki bilgilere ulaşılabilir:

- 1- MYT'deki ölçme aracı ve madde parametreleri ile  $\theta$  düzeyleri için var olan değişmezlik özelliği, KTT'de geçerli değildir.
- 2- KTT'de, ölçme sonuçları  $\theta$ 'nın farklı gruplarındaki bireyler için aynı güvenilirlikle tahmin edilirken; MYT'de ölçme aracının ne derece güvenilir olduğu  $\theta$ 'nın her düzeyi için ayrı ayrı elde edilir.
- 3- KTT'de ölçme aracının güvenilirliği, madde sayısının bir fonksiyonudur. Ölçme aracındaki madde sayısı arttıkça, güvenilirlik de artar. MYT'de, daha kısa ama benzer düzeyde güvenilirliğe sahip ölçme araçları, uygun madde kullanımı ile geliştirilebilir.
- 4- Madde seçiminde, KTT temel olarak madde ayırt ediciliğini, madde varyansını, maddeler ve maddelerle toplam puanlar arasındaki ilişki katsayılarını kullanırken; MYT madde parametrelerini, madde bilgi fonksiyonlarını, model-veri uyum istatistiklerini kullanır.
- 5- KTT, bir bireyin bir ölçme aracında nasıl bir performans göstereceği hakkında önceden bilgi vermez. MYT'de ise;  $\theta$  düzeyi bilinen bir bireyin, parametreleri bilinen herhangi bir maddedeki performansı tahmin edilebilir.
- 6- MYT hem iki sonuçlu hem de çok sonuçlu puanlanan maddelerde kullanılabilecek modellere sahiptir.
- 7- MYT modellerinin önemli bir özelliği, madde ve ölçme aracının işlevinin grafiksel gösterimleridir. MYT modelleri, bireyin  $\theta$  düzeyi verildiğinde, bu bireyin maddedeki performansını,  $\theta$ 'nın tüm düzeylerinde ölçme aracının güvenilirliğini/kesinliğini ve maddelere yanıt verme bakımından grup farklılıklarını grafiksel olarak göstermeye olanak sağlar. Bu grafiksel gösterimler, sonuçların sunumu ve MYT hakkında fazla bilgi sahibi olmayanların teorisinin yöntem bilimini daha iyi anlamalarını sağlama bakımından oldukça değerlidir (Hambleton ve Swaminathan, 1985).

## 1.5. Bilgisayar Uyarlamalı Test Yöntemi

### 1.5.1. Bilgisayar Uyarlamalı Test Yöntemine Giriş

“Uyarlamalı test” yöntemi, bir ölçme aracını oluşturan maddelerin, uygulama sırasında seçildiği bir süreci tanımlamak için kullanılan çok sayıda terimden birisidir (Ware ve ark., 2003). Kapsam ve madde sayısı bakımından, bir ölçme aracında yer alan maddelerin tümü her bireye uygulanırsa, ölçme aracı “sabit ve değişmez” olarak tanımlanırken; madde sayısı değişkenlik gösterdiğinde ve/veya maddeler her bir birey için farklılık gösterdiğinde, “uyarlamalı” kategorisinde değerlendirilir (Blais ve Raiche, 2002). Uyarlamalı test ile dönüşümlü olarak kullanılan diğer terimler; adapte edilen test yöntemi/bireye göre ayarlanan test yöntemi (tailored testing), bölümlendirilmiş test yöntemi (branched testing), programlanmış test yöntemi (programmed testing), bireyselleştirilmiş test yöntemi (individualized testing)’dir. Ölçme aracı bireye göre uyarlandığı için, uyarlamalı terimi daha yaygın olarak kullanılmaktadır (Ware ve ark., 2003).

BUT, 1905 yılında Binet ile başlayan uyarlamalı test uygulamalarının bir serisidir. Uyarlamalı testler, iteratif bir süreç olup; ölçme aracından seçilen maddeler kullanılarak gerçekleştirilir. Maddeler, bireyin  $\theta$  düzeyi ile eşleşecek şekilde seçilir. Bu süreçte ilk olarak, soru bankasındaki maddelerden orta derecede zorluğa sahip olan biri seçilir ve birey yanıt verir. Verilen yanıtı göre bireyin  $\theta$  düzeyine ilişkin tahmin elde edildikten sonra, soru bankasındaki hangi maddenin bireyin  $\theta$  tahmini için en fazla bilgi sağlayacağına karar verilir. Maddelere verilen yanıtlara dayalı olarak, yeni bir  $\theta$  tahmini hesaplanır. Bireyin  $\theta$  düzeyi istenen doğruluk düzeyinde belirlendiğinde test sona erer.

### 1.5.2. Bilgisayar Uyarlamalı Test Yönteminin Tarihsel Gelişimi

Uyarlamalı testler, temel olarak uygulayıcıların gereksinimlerini karşılamak ve bireylerin  $\theta$  düzeylerini belirlemek için oluşturulmuştur. Bireylere çok kolay bir ölçme aracı uygulamak, zaman kaybının yanı sıra dikkatsizce hata yapılması ve kritik maddelere bilerek yanlış yanıtların seçilmesi gibi istenmeyen birey davranışlarına neden olabilmektedir. Diğer taraftan çok zor bir ölçme aracı uygulamak, bilgi verici olmayan sonuçları ortaya çıkarmaktadır. Çünkü bireyler maddeleri yanıtlamak için ciddi bir girişimde bulunmayacaklar; tahmin yoluyla, yanıt seçeneklerinden yararlanarak ya da istenmeyen davranışlar göstererek doğru yanıt bulmaya çalışacaklardır. Bununla beraber ölçme aracının her bir bireyin  $\theta$  düzeyini karşılayacak şekilde uyarlanması ve sonuçta bazı bireylere daha az, diğerlerine daha çok madde uygulanmasından dolayı uyarlamalı testler adaletsiz olarak değerlendirilebilmektedir. Bu durumda ortaya “Her bir bireye farklı bir test uygulama yoluyla bireyler nasıl karşılaştırılabilir?” sorusu çıkmaktadır.

Alfred Binet, bu alandaki en büyük gelişmeyi 1905 yılında kendi geliştirdiği zekâ testleri (Binet’s IQ tests) ile gerçekleştirmiştir. İlgilendiği durum, gruptan daha çok her bir bireyin  $\theta$  düzeyinin belirlenmesi olduğu için, bireylerin tümünün aynı maddeleri yanıtlanmasını gerektirecek bir zorunluluğun olmadığını belirtmiştir. Ölçme aracının, her bir bireye basit bir strateji yardımıyla uyarlanabileceğini, bireyin  $\theta$  düzeyine uygun olan maddenin uygulanması ile test sürecinin başlayabileceğini ortaya koymuştur. Bu yaklaşımda, birey doğru yanıt verirse, yanlış yanıt verene kadar daha zor maddeler, yanlış yanıt verirse, doğru yanıt verene kadar daha kolay maddeler uygulanmakta, test sonucunda bireyin  $\theta$  düzeyi tahmin edilmektedir. Binet’in yaklaşımı, bilgisayar ile uygulanması kolay bir süreçtir. Günümüzde Stanford-Binet testleri ve çocuklar ile yetişkinler için kullanılan ve daha popüler olan Wechsler zeka testleri, maddelerin seçiminde uyarlamalı yaklaşımı kullanmaktadır (Microsoft Inc., 2006). Lord’un esnek



düzeyle test (flexilevel testing) yaklaşımı ve bunun çeşitleri, Henning'in aşama (step) yaklaşımı ve Lewis ile Sheehan'ın test kitapçıkları (testlets), Binet'in yaklaşımının daha basitleştirilmiş halleri olup, klasik yöntem ya da bilgisayar yardımıyla uygulanabilir (Linacre, 2000).

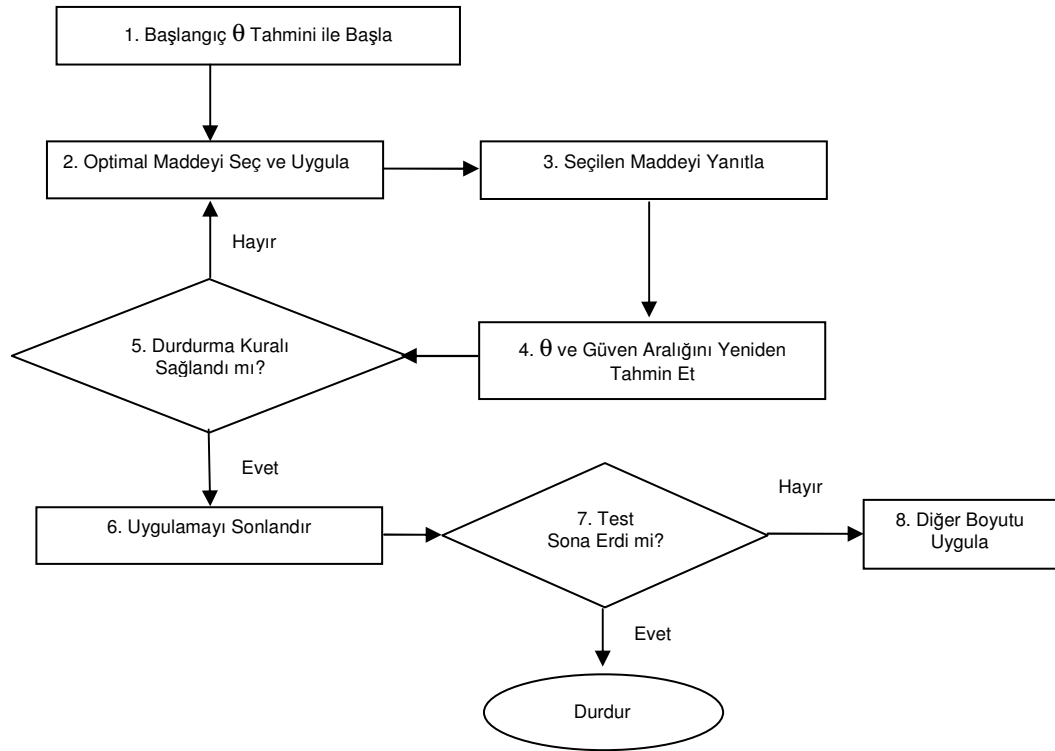
Gerçekte, uyarlamalı testler için kullanılacak model ve stratejiler, hemen hemen 50 yıldır var olmasına rağmen; bilgisayar donanım ve yazılımlarının nadir, pahalı ve kullanışsız olması, BUT yönteminin uygulanmasını sınırlandırmıştır (Blais ve Raiche, 2002). Günümüzde bilgisayarların, test uygulamada yaygın kullanımı ile uyarlamalı testler daha popüler olmuştur. Bilgisayar, bireyin  $\theta$  düzeyinin tahmin edilmesi için gerekli hesaplamaları yapabilmekte ve birey için en uygun maddeleri seçebilmektedir. Yaygın olarak kullanılan bazı testlerin, uyarlamalı olarak adaptasyonları yapılmıştır. Dünyanın en büyük test organizasyonu olan "The Educational Testing Service (ETS)", GRE sınavını 1993 yılında uyarlamalı test olarak yayınlamıştır. TOEFL sınavı da, 2000 yılında BUT yöntemine uygun hale getirilmiştir. Bilgi teknolojileri endüstrisinde, Novell firması 1991 yılında sertifika programlarında BUT yöntemini kullanmaya başlamıştır. Aktif olarak kullanılan test programlarına ek olarak, geliştirme aşamasında olan çok sayıda pilot program da bulunmaktadır. Bunların en yaygın olanları ACT (American College Test), SAT (Scholastic Aptitude Tests and Scholastic Assessment Tests) ve ASVAB (Armed Services Vocational Aptitude Battery)'dir (Scantron Corp., 2003; Microsoft Inc., 2006).

### **1.5.3. Bilgisayar Uyarlamalı Test Yönteminin İşleyişi**

Temel olarak, BUT yönteminin uygulama süreçleri oldukça basit ve açıktır. BUT, ölçme aracından seçilen maddeler kullanılarak gerçekleştirilir. BUT yönteminde bireyin ilk maddeye verdiği yanıt, sonraki maddelerin sıralamasının belirlenmesinde önemli rol oynar. Eğer ilk madde doğru yanıtlandıysa, sonraki madde daha zor, yanlış yanıtlandıysa sonraki madde

daha kolay olacaktır. Bu yaklaşımın arkasındaki mantık, çok kolay ya da çok zor maddelerden bireyin  $\theta$  düzeyi hakkında bilgi edinilemeyeceği, dolayısıyla maddelerin bireyin  $\theta$  düzeyi hakkında en fazla bilgiyi sağlayacak şekilde seçilmiştir. Dolayısıyla birey, zor maddeden daha fazla puan alırken, kolay maddeden daha düşük puan alır. BUT ile bireyin  $\theta$  düzeyi, test süreci boyunca iteratif olarak tahmin edilir ve maddeler elde edilen tahminlere dayalı olarak seçilir. Uyarlamalı madde seçimi, seçilen maddelerin bireyin  $\theta$  düzeyine uygun olması nedeniyle düşük  $SE(\theta)$  ve yüksek kesinlik değerinin elde edilmesine olanak sağlar (Rudner, 1998).

BUT yönteminin mantığı Şekil 1.5'de gösterilmiştir (Wainer ve ark., 2000).



Şekil 1.5: BUT Yönteminin Mantığı

Test, bireyin  $\theta$  düzeyine ilişkin başlangıç tahmini ile başlar. Bu tahmin, tüm bireylere sorulan genel bir maddeye verilen yanıtta ya da birey hakkında önceden elde edilen bilgiye dayalı olabilir.

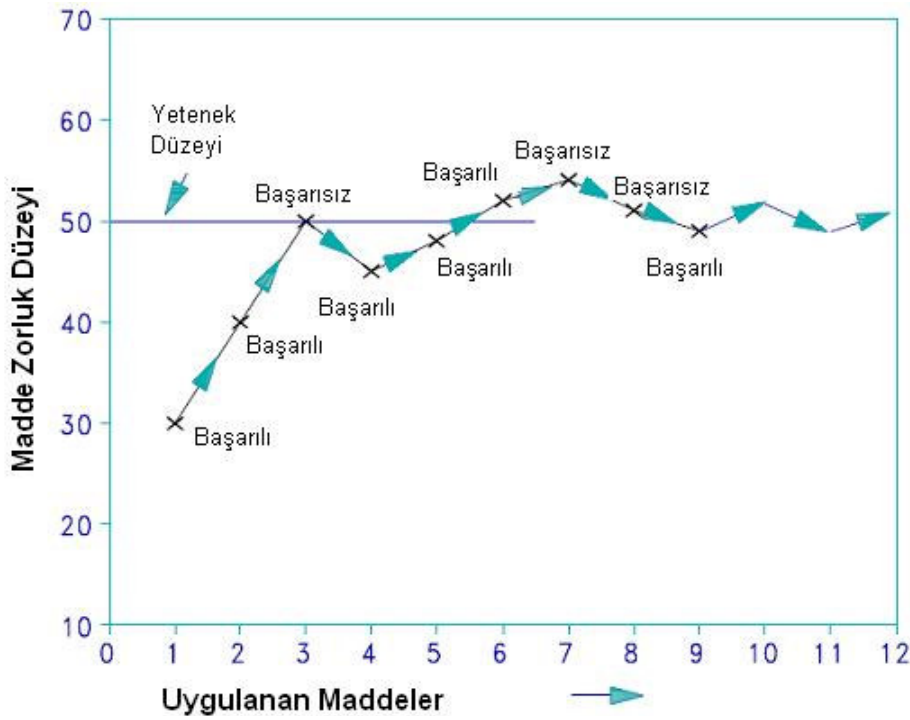
İkinci aşamada henüz uygulanmamış olan tüm maddeler,  $\theta$  düzeyine ilişkin başlangıç tahmin değeri verildiğinde, hangi maddenin sıradaki en iyi madde olacağına karar vermek için değerlendirilir. Maksimum bilgi yaklaşımı altında, sıradaki en iyi madde,  $\theta$  düzeyinde en fazla bilgiyi sağlayandır. Bu yüzden, 2.aşamada, her bir madde için  $I_i(\theta)$ , en güncel  $\theta$  tahmini kullanılarak değerlendirilebilir. Bilginin maksimize edilmesi, maddelerin seçilmesinde belki de en iyi bilinen yaklaşım olmasına rağmen, Kingsburg ve Zara (1989), madde seçimi için çeşitli alternatif süreçler tanımlamışlardır.

İkinci aşamada seçilen optimal maddeye, üçüncü aşamada verilen yanıt, 4. aşamadaki  $\theta$  düzeyini ve güven aralığını yeniden tahmin etmek için kullanılır. Bu aşamada verilen yanıt yapısı için, en çok olabilirlik yaklaşımı ya da Bayesçi yaklaşımlar kullanılarak  $\theta$  düzeyine ilişkin tahmin değeri güncellenir.

Beşinci aşamada, önceden belirlenmiş durdurma kuralının sağlanıp sağlanmadığı kontrol edilir. Durdurma kuralı; zaman, uygulanan madde sayısı,  $\theta$  düzeyine ilişkin tahmin değerindeki değişim, standart hata gibi bir kesinlik göstergesi ya da bu faktörlerin bir kombinasyonu olabilir. Kural bir kere sağlandığında, bilgisayar diğer boyutu değerlendirmeye başlar ya da testi sonlandırır. Bu durum gerçekleşmezse, süreç Aşama 2'den başlayarak 4.aşamada elde edilen  $\theta$  düzeyi için en çok bilgi sağlayan madde ile devam eder (Scantron Corp., 2003).

### 1.5.3.1. Bilgisayar Uyarlamalı Test Yönteminin İki Sonuçlu Maddelerde İşleyişi

BUT'un iki sonuçlu maddeler için nasıl çalıştığını göstermek amacıyla Şekil 1.6'da verilen örnek açıklanmaya çalışılacaktır. Çalışmadaki ölçme aracının iki sonuçlu maddelerden oluştuğunu ve her bir maddenin, 0 ile 100 aralığında değişen bir b değerine sahip olduğunu varsayalım. Bu koşullar altında Şekil 1.6'daki süreci inceleyelim.



Şekil 1.6: BUT Yönteminin İki Sonuçlu Maddelerde İşleyişi

- Uygulanan ilk maddenin b değerinin 30 birim ve bireyin  $\theta$  düzeyinin 50 birim olduğunu varsaydığımızda, bireyin ilk maddeyi büyük olasılıkla doğru olarak yanıtlaması beklenir.
- İkinci madde, ilk maddeden biraz daha zor olup; b değeri 40 birim olsun ve birey bu maddeyi doğru yanıtlasın.
- Bilgisayar, b değeri 50 birim olan daha zor bir madde seçsin. Bireyin  $\theta$  düzeyi ve maddenin b değeri birbirine eşit olduğundan bireyin, bu

maddede başarılı olma olasılığı %50'dir. Birey bu maddeye yanlış yanıt versin.

- Bu aşamada, b değeri 50 birimden daha düşük, fakat 40 birimden bir miktar daha yüksek bir madde uygulansın. Örneğe göre, b değeri 45 birim olan bir madde seçilsin ve birey doğru yanıt versin.
- Beşinci maddenin b değeri 48 birim olsun ve birey bu maddeye de doğru yanıt versin.
- Bilgisayar, b değeri 48 birimden biraz daha yüksek olan bir madde uygulansın ve birey doğru yanıt versin.
- Bireyin, b değeri 40 ve 48 birim olan maddelere verdiği yanıtlar incelendiğinde, b değeri 50 birim olan bir maddeye yanlış yanıt vermesinin şanssızlık olabileceği düşüncesi ile bilgisayar, b değeri 52 birim olan bir maddeyi uygulansın ve birey doğru yanıt versin.
- Bu aşamada bilgisayar, b değeri 54 birim olan bir madde seçsin ve birey yanlış yanıt versin.
- Bilgisayar tarafından seçilen maddenin b değeri 51 birim olsun ve birey yine yanlış yanıt versin.
- Bilgisayar, b değeri 49 birim olan bir madde seçsin ve birey doğru yanıt versin.

Bu süreç, seçilen durdurma kuralı sağlanana kadar devam eder. Dikkat edilirse, bir süre sonra bireyin  $\theta$  düzeyi, madde zorluk parametresi 50 birim etrafında değişmeye başlamaktadır (Linacre, 2000).

### **1.5.3.2. Bilgisayar Uyarlamalı Test Yönteminin Çok Sonuçlu Maddelerde İşleyişi**

Temel olarak, BUT yönteminin çok sonuçlu maddelerde uygulanması, iki sonuçlu maddelerde uygulanması ile aynı olup; farklılık yanıtların skorlanmasıdır. Doğru yanıt en yüksek skora sahip iken; bazı çeldiriciler,



eşik değeri 35 olduğu için, bu bireyin  $\theta$  düzeyinin 35'den daha yüksek olacağı tahmin edilir.

- İlk maddenin, birey için kolay olduğu kanıtlandığından, ikinci maddenin biraz zor olması hedeflensin. Bu madde için 1 yanıtını vermesi beklenen birey, 2 yanıtı versin.
- Bu durumda üçüncü madde daha zor seçilsin ve birey tekrar 2 yanıtını versin.
- Dördüncü madde ile bireyin kesin başarı sağlayabileceği en yüksek düzeyin ne olduğu bulunmaya çalışılır. Bireyin en yüksek kategoriye seçebileceği daha kolay bir madde seçilsin ve birey 1 yanıtı versin.
- Beşinci madde için bireyin en yüksek kategoriye seçebileceği düzeyin ne olduğunun bulunması yönünde yeni bir madde belirlensin ve birey 3 yanıtı versin.
- Bireyin tamamen başarısız olmasından önce bir maddenin ne derece zor olması gerektiğini belirlemek için biraz daha zor olan 6. madde seçilsin ve birey 1 yanıtı versin.
- Bu aşamada, 6. maddeden daha kolay, fakat yine de zor olan 7. madde seçilsin ve birey 2 yanıtı versin.
- Bu durumda biraz daha zor bir madde olan 8. madde seçilsin ve birey 0 yanıtı versin.

Bu süreç, seçilen durdurma kuralı sağlanana kadar devam eder. Çok sonuçlu maddeler, bireyin  $\theta$  düzeyi hakkında daha fazla bilgi verici olduğu için, çok sonuçlu maddeler ile yapılan BUT uygulamaları genellikle daha az sayıda madde ile sona erer (Linacre, 2000).

## 1.5.4. Bilgisayar Uyarlamalı Test Yönteminde Önemli Noktalar

### 1.5.4.1. İlk Maddenin Seçimi

İlk maddenin seçimi, seçilen yaklaşımın en çok olabilirlik ya da Bayesçi olmasına bağlıdır. Eğer en çok olabilirlik yaklaşımı kullanılıyorsa ilk madde, ortalama/ortanca zorluk değerine sahip olan maddedir. Kullanılan yaklaşım Bayesçi yaklaşım ise, başlangıç  $\theta$  düzeyi, önsel  $\theta$  dağılımının ortalamasına eşit olacak şekilde belirlenir. Bu durumda ilk madde başlangıç  $\theta$  düzeyinde fazla/en fazla bilgi sağlayan madde olacaktır (Segall, 2003).

### 1.5.4.2. Madde Seçimi

Uyarlamalı yaklaşımda, madde seçimi için yaygın olarak kullanılan iki yaklaşım vardır (Kingsbury ve Zara, 1989). İlki “maksimum bilgi” yaklaşımı (Weiss, 1982) olup;  $\theta$  düzeyinde en fazla bilgiyi sağlayan maddenin seçimi temeline dayalıdır. Aynı maddelerin tekrar seçilmesini engellemek ve test güvenliği ile test geçerliğini etkilememek için Green ve ark. (1984), maddelerin  $\theta$  düzeyinde en fazla bilgiyi sağlayan maddeler arasından rasgele bir temele dayalı olarak seçilmesini önermişlerdir.

İkinci yöntem, “Bayesçi” yaklaşım (Owen, 1975) olup; bireyin  $\theta$  düzeyinin sonsal dağılımının varyansını minimize eden ölçme aracı maddelerinin seçimi temeline dayalıdır. Bayesçi yaklaşımda, bireyin  $\theta$  düzeyi hakkında önsel bir varsayımdan yola çıkılarak, sonsal  $\theta$  düzeyi belirlenmeye çalışılır. Bu nedenle yaklaşımın başarısı, seçilen önsel dağılıma bağlıdır (Hambleton ve ark., 1991).



### 1.5.4.3. Durdurma Kuralları

Bireye uygulanan testin hangi koşulda sona ereceği kararı, BUT'da en temel unsurlardan birisidir. Eğer test çok kısa ise, bireye ilişkin  $\theta$  tahmini hatalı olabilirken; çok uzun olduğunda, bireylerde bıkkınlık ve performans düzeylerinde düşüş gözlenebilmektedir. BUT'da iki temel durdurma yaklaşımı vardır: sabit uzunluk ve değişken uzunluk.

Sabit uzunluğa sahip testler, her bir bireyin aynı sayıda maddeye yanıt vermesini gerektirir. Sabit uzunluklu testlerde,  $\theta$ 'ya ilişkin kesinlik, bireyler arasında değişkenlik gösterir. Bu durumun tersine, değişken uzunluğa sahip testler, belirlenen bir kesinlik düzeyine, maksimum madde sayısına (bu değer genellikle klasik yöntemdeki madde sayısına eşittir) ya da her bir birey için belirlenen maksimum test süresine ulaşılan kadar devam eder. Bu yaklaşımdaki amaç, bireye uygulanan her bir madde ile  $\theta$  düzeyine ilişkin tahminin  $SE(\theta)$  değerinin azaltılarak, tahmin değerinin kesinliğinin artırılmaya çalışılmasıdır (Segall, 2003).

### 1.5.4.4. Soru Bankası

Soru bankası, BUT yönteminin etkinliği üzerinde birinci derecede öneme sahiptir. Uygun soru bankası ile BUT, klasik yöntemden daha etkili olabilir. BUT yönteminin uygulanabilmesi için gerekli olan, madde parametreleri bilinen, çok sayıda maddeden oluşan bir soru bankasına sahip olmaktır. Test uygulamasının amacı,  $\theta$ 'nın tüm aralığında iyi ölçüm yapmak ise, ideal soru bankası,  $\theta$ 'nın geniş aralığında yayılım gösteren yeterli sayıda maddeden oluşmalıdır. Eğer amaç, bireyin belirli bir kesim noktasının altında ya da üstünde olup olmadığına ilişkin karar vermek ise, kesim noktası civarında b değerine sahip olan maddelerden oluşan bir soru bankası uygundur (Embretson ve Reise, 2000).

BUT uygulanmadan önce soru bankasının (bankalarının) psikometrik özellikleri değerlendirilmelidir. Soru bankasında bulunması istenen en önemli özellikler; soru bankasının tek boyutlu bir yapıya sahip olması, soru bankasındaki maddelerin incelenen MYT modeline uyum göstermesi ve yaş, cinsiyet, eğitim durumu gibi faktörlerden bağımsız olarak çalışması (MIF göstermemesi) gerekmektedir.

### 1.5.5. Bilgisayar Uyarlamalı Test Yönteminin Avantajları

BUT yönteminin avantajları şunlardır (Rudner, 1998):

- Yeterli kesinliğe erişmek için daha az sayıda maddeye gereksinim olduğu için BUT'ta, sabit sayıda maddeden oluşan klasik yöntemle göre daha az zamana gereksinim vardır. Ölçme araçları, aynı düzeyde güvenilirliği sağlayarak, uygulama zamanında yaklaşık %50'lik bir azalmaya neden olurlar. Daha kısa uygulama süresi, bireyin sonuçlarını etkileyebilecek bir faktör olan bıkkınlığı azaltır.
- BUT uygulaması kalibre edilmiş maddelerden oluşan bir soru bankası ile yapıldığından, değerlendirme rutinde kullanılan ölçme araçlarına göre daha yüksek kesinlikte sonuç verir.
- Ölçme araçları, genellikle  $\theta$ 'nın orta düzeyindeki bireyler için daha doğru sonuçlar verirken; BUT  $\theta$ 'nın geniş bir aralığında doğru sonuçlar verir.
- Değerlendirme istendiği anda uygulanır ve sonuçlar anında belirlenir.
- Çok merkezli yapılan çalışmalarda verilerin tek bir merkezde toplanmasına olanak sağlayarak, çalışmanın standartlaştırılmasını kolaylaştırır.
- Ölçme aracına ilişkin kâğıtlar bireylere dağıtılmadığı için, gizlilik açısından test güvenliği daha yüksektir.

### 1.5.6. Bilgisayar Uyarlamalı Test Yönteminin Dezavantajları

BUT yönteminin dezavantajları şunlardır (Rudner, 1998):

- Donanım sınırlılıkları, bilgisayar yoluyla uygulanabilen madde türlerini kısıtlar. Detaylı çizim içeren maddelerin BUT yönteminde gösterimi zor olabilir.
- Klasik yöntemle elde edilen madde parametreleri, BUT yönteminde kullanılamaz.
- Bilgisayar uyarlamalı versiyonlarla, bireyin metinlerin altını çizmesi veya elenen yanıtların üzerini çizmesi mümkün değildir. Yapılan çalışmalar, bilgisayar ekranlarının, basılı materyallerin okunması bakımından daha fazla zaman aldığını göstermiştir.

## 2. GEREÇ VE YÖNTEM

Bu çalışmanın amacı, toplumda en sık rastlanılan iki temel kas-iskelet sistemi sorununda (bel ağrısı ve diz osteoartriti) özürülük değerlendirimi için soru bankaları geliştirilerek BUT yönteminin birer uygulamasının yapılmasıdır.

Çalışma kapsamında öncelikli olarak soru bankasında yer alması düşünülen maddeler, bel ağrılı ve diz osteoartritli hastaların değerlendirmesinde kullanılan, kas-iskelet sorunlarına spesifik olarak “bozukluk”, “aktivite sınırlanması”, “katılım kısıtlanması” ve “yaşam kalitesi”ni değerlendiren mevcut ölçeklerin incelenmesi sonucunda belirlenmiştir. Bu seçimde temel alınan ICF kategorileri bel ağrısı ve diz osteoartrite özgün daha önce belirlenmiş olan ICF çekirdek dizinini içermiştir (Cieza ve ark., 2004; Dreinhofer ve ark., 2004). Bu inceleme sonucunda bel ağrılı hastalar için WHODAS-II, Oswestry, Roland Morris ve NHP; diz osteoartritli hastalar için WHODAS-II, WOMAC, HAQ ve NHP ölçeklerinin kullanılması uygun bulunmuştur. Çalışmada kullanılan ölçekler hakkında özet bilgi aşağıda sunulmuştur.

**WHODAS-II (World Health Organization-Disability Assessment Schedule II):** Tıbbi tanıdan bağımsız olarak, bireyin faaliyet düzeyinde ve topluma katılımında yaşadığı kısıtlılıkları tespit edebilmek amacıyla geliştirilen ölçekte 36 madde yer almaktadır. Birçok kültürde ortak olarak önemli sayılan faaliyetleri içeren 6 alandan (anlama ve iletişim kurma, hareket etme-bir yerden bir yere gitme, kendine bakım, insan ilişkileri, yaşam faaliyetleri, toplumsal yaşama katılım) oluşmaktadır. Tüm bu alanlarla ilgili maddelerde kişinin son bir ayda o faaliyet sırasında ne kadar güçlük çektiği sorgulanmakta, yanıtlar 1–5 arasında (1: hiç, 5: aşırı/hiç yapamıyorum) puanlanmaktadır.

**Oswestry Bel Ağrısı Özürülük Sorgulaması:** Ağrı düzeyini, çeşitli fiziksel aktiviteleri, kendine bakım aktivitelerini, uykuyu, sosyal hayatı, seyahati, cinsel yaşamı sorgulayan ölçek, 10 maddeden oluşmaktadır ve her madde 0–5 arasında (0: normal, 5: maksimum zorluk) puanlanmaktadır.

**Roland-Morris Bel Ağrısı Özürülük Sorgulaması:** 24 maddeden oluşan, her maddenin evet/hayır şeklinde yanıtlandığı bel ağrısına bağlı özürülülüğü değerlendiren bir ölçektir.

**NHP (Nottingham Health Profile):** Kişinin kendisinin algıladığı sağlık durumunu fiziksel, emosyonel ve sosyal açılardan ölçmeyi amaçlayan jenerik bir yaşam kalitesi ölçeğidir. Her biri evet/hayır şeklinde yanıtlanan toplam 38 maddeden oluşmaktadır.

**WOMAC Osteoartrit İndeksi [Western Ontario and McMaster Universities (WOMAC) Osteoarthritis Index]:** 24 maddeden oluşan, ağrı, tutukluk ve fiziksel özürülük olmak üzere üç boyutu irdeleyen bir ölçektir. Her madde 0–4 arasında (0: yok, 4: çok şiddetli) puanlanmaktadır.

**HAQ (Health Assessment Questionnaire):** HAQ, 8 bölüm (giyinip-kuşanma, doğrulma, yemek yeme, yürüme, hijyen, uzanma, kavrama ve günlük işler) şeklinde 20 maddeden oluşan, günlük yaşam aktivitelerini değerlendiren bir fiziksel özürülük ölçeğidir. Her madde 0–3 arası (0: hiç zorluk çekmeden yapıyorum, 3: hiç yapamıyorum) puanlanmaktadır (Küçükdeveci, 2004).

Başlangıç soru bankasını belirleme işleminden sonra, çalışmada yer alacak bireyler, Ankara Üniversitesi Tıp Fakültesi Fiziksel Tıp ve Rehabilitasyon (FTR) Anabilim Dalı kliniğinde bel ağrısı ve diz osteoartriti tanısı ile tetkik ve tedavi edilen hastalardan gönüllü olanlar arasından, olurları alınarak çalışmaya dahil edilmiştir. Seçilen 266 bel ağrılı ve 225 diz

osteoartritli hasta, aynı anabilim dalında ve Biyoistatistik anabilim dalında görev yapan araştırma görevlileri tarafından, başlangıç soru bankasını oluşturan maddeler kullanılarak yüz yüze görüşme yöntemiyle (klasik kağıt-kalem testi) değerlendirilmiştir.

BUT için kullanılacak final soru bankasının belirlenmesi, klasik test sonucunda elde edilen veriler kullanılarak üç aşamada yapılmıştır.

### **Aşama I: Boyutların Belirlenmesi**

İlk aşamada her bir hastalık grubu için ayrı ayrı oluşturulan başlangıç soru bankalarının tek boyutluluk varsayımını sağlayıp sağlamadığını belirlemek için kategorik verilerde açıklayıcı ve doğrulayıcı faktör analizi (AFA ve DFA) uygulanmıştır. AFA ile maddelerin kaç boyutta açıklandığı; yapısal eşitlik modellenmesine dayalı DFA ile de her bir faktör tarafından o faktörleri oluşturan maddelerin kovaryans yapısını açıklayıp açıklayamadığı incelenmiştir. Bu analizler için **MPlus** (Muthen ve Muthen, 2006) paket programı kullanılmıştır.

AFA, temel bir madde setinden yola çıkarak; herhangi bir önsel bilgi olmaksızın bu maddelerden oluşturulacak gizli yapı sayısının belirlenmesinde kullanılan istatistiksel bir yöntemdir. Geleneksel faktör analizi yöntemi, olduğundan daha fazla sayıda boyut belirlediği için ve temel bileşenler analizi (TBA) temeline dayandığı için kategorik verilerin analizinde kullanımı uygun değildir. Bu nedenle kategorik verilerin analizinde polikorik ve tetrakorik korelasyonları temel alan ağırlıklandırılmış en küçük kareler yöntemi kullanılmıştır (Suhr, 2006).

DFA, başlangıç soru bankasında yer alan maddelerden elde edilen faktör yapılarını doğrulamak için Yapısal Eşitlik Modellerini (YEM) kullanarak gerçekleştirilir. Bu analizde temel hipotez, "maddeler ile ait oldukları faktörler

arasında bir ilişki vardır” biçiminde kurulur. Daha sonra, teorik bilgi, deneysel araştırma ya da her ikisi de kullanılarak ilişki yapısına ilişkin önsel bilgi ışığında yukarıdaki hipotez istatistiksel olarak test edilir. Klasik istatistiksel yöntemler, analizin önemliliğini belirlemek için tek bir istatistiksel test kullanırken; DFA verinin modele uyumunu belirlemek için çeşitli istatistiksel testlerden yararlanır. Modelin uyum istatistikleri yardımıyla faktörlerin gerçekten bu maddelerden oluşup oluşmadığına karar verilir. Sonuç olarak YEM’e dayalı DFA’nın amacı, önceden belirlenen bu ilişki örüntülerinin veri tarafından doğrulanıp doğrulanmadığını ortaya koymaktır.

Başlangıç soru bankasında yer alan maddelerin kaç boyutta toplanması gerektiğine karar verirken uyum istatistiklerinden yararlanılmıştır. Uyum istatistikleri, modelin kabul edilip edilemeyeceğine ilişkin sınır değerler kullanılarak yorumlanır. En yaygın kullanılan uyum iyiliği istatistikleri, uyum iyiliği indeksi (Uİ) [*Goodness of Fit Index-(GFI)*], düzeltilmiş uyum iyiliği indeksi (DUİ) [*Adjusted Goodness of Fit Index (AGFI)*], karşılaştırmalı uyum indeksi (KUI) [*Comparative Fit Index (CFI)*], Tucker-Lewis indeksi (TLİ) [*Tucker-Lewis Index (TLI)*], Yaklaşımın Hata Kareler Ortalaması Karekökü (YHKOK) [*Root Mean Square Error of Approximation (RMSEA)*], Artık Kareler Ortalaması Karekökü (AKOK) [*Root Mean Square Residual (RMSR)*] ve Ağırlıklandırılmış Ortalama Artıkların Karekökü (AOAK) [*Weighted Root Mean Residuals (WRMR)*]’dir. Bunların ilk dördünün değerlerinin 0.90’dan büyük olması, kabul edilebilir bir uyum, 0.95’den büyük olmaları ise iyi bir uyum olduğunun göstergesi olarak kabul edilir. YHKOK, AKOK ve AOAK’da ise söz konusu değerlerin 0.05’in altında olması iyi bir uyum, 0.08’in altında olması ise kabul edilebilir bir uyum olduğunu gösterir (Şimşek, 2007).

AFA ve DFA sonucunda birden fazla boyut belirlendiğinde, soru bankası bu alt boyutlar için ayrı ayrı oluşturulmuştur. Her bir boyutta yer alan maddelerden soru bankaları oluşturulurken, bu maddelerin ICF sınıflamaları göz önünde bulundurulmuş ve elde edilen boyutların mantıklı bir şekilde isimlendirilmesi gerçekleştirilmiştir. Faktör analizi sonucunda herhangi bir

boyuta yüklenmeyen (her bir boyuttaki faktör yükü  $<0.40$  ve farklı faktörlere ilişkin yükler arasındaki fark  $<0.10$  olan) maddeler çalışmadan çıkarılmıştır. Bu aşamada amaç soru bankası oluşturmak olduğundan, eğer bir madde birden fazla boyuta yükleniyorsa (her bir boyuttaki faktör yükü  $>0.40$ ), anlamlı faktör yüküne sahip olduğu boyutların her birinde de yer almıştır (Fliege ve ark., 2005).

## **Aşama II: Psikometrik Özelliklerin İncelenmesi**

Sağlık alanında hastaların  $\theta$  düzeylerinin belirlenmesi için geliştirilen ölçme araçları, geçerli değişim skorlarının ve parametrik istatistiklerin elde edilmesinin gerekli olduğu klinik denemelerde yaygın olarak kullanılır. Rasch analizi, veri seti Rasch modelinin beklentilerine uyum sağladığı durumda sıralı ölçekle elde edilmiş yanıtların, aralık değişkenine dönüşümünü sağlayarak, hastaların doğru bir şekilde değerlendirilmesine yardımcı olur. Rasch analizi farklı durumlarda kullanım alanına sahiptir. Bunlar; yeni bir ölçeğin geliştirilmesi, var olan bir ölçme aracının psikometrik özelliklerinin değerlendirilmesi, BUT için soru bankalarının oluşturulmasıdır. Bu alanların her biri için aşağıdaki standart protokolün yerine getirilmesi gerekmektedir:

- Tek boyutluluk için ölçme aracının iç tutarlılığının test edilmesi
- Ölçme aracındaki maddelerin modele uyumunun incelenmesi
- Maddelerin değişmezlik özelliğini sağlayıp sağlamadığının test edilmesi
- Maddelere ilişkin eşik değerlerinin sıralı olup olmadığının test edilmesi
- MIF'in test edilmesi

(Pallant ve Tennant, 2007; Tennant ve Conaghan, 2007).



Çalışmanın bu aşamasında, birinci aşamada elde edilen soru bankalarının psikometrik özellikleri **RUMM 2020** (RUMM Laboratory Pty. Ltd., 2007) paket programı yardımıyla yukarıda bahsedilen protokol uygulanarak incelenmiştir. Soru bankalarının psikometrik özelliklerini değerlendirmede, MYT altında bulunan Rasch modellerinin kullanımı standart hale gelmiştir. Çünkü Rasch modeline uyum sağlayan ölçme araçlarının psikometrik özelliklerinin yeterli olduğu kabul edilmektedir (Tennant ve Conaghan, 2007).

Çalışmamızda psikometrik özelliklerin incelenmesinde Rasch modellerinden kısmi kredi modeli kullanılmıştır. Kısmi kredi modeli, eşit ayırt edicilik gücüne sahip olduğu varsayılan çok sonuçlu maddeler için uygun bir modeldir. *i*. maddeye verilen yanıtlar, sırayla tamamlanması gereken ( $m_i+1$ ) tane sıralı kategori olarak sınıflandırılır. *i*. ( $i=1,2,\dots,n$ ) madde için kategoriler ( $k$ ),  $0,1,\dots,m_i$  değerlerini alabilen birbirini izleyen tamsayı değerlerdir ve başarı ile tamamlanan aşamaların sayısını gösterir. Belirli bir  $\theta$  değerine sahip olan bir bireyin, *i*. maddede  $k$  kategorisini seçme olasılığı aşağıdaki gibidir:

$$P_{ik} = \frac{\exp \sum_{c=0}^k (\theta - b_{ic})}{\sum_{c=0}^{m_i} \left[ \exp \sum_{j=0}^c (\theta - b_{ij}) \right]} \quad (\text{Denklem 2.1})$$

$$b_{i0} \equiv 0, \quad \sum_{c=0}^0 (\theta - b_{ic}) = 0 \quad \text{ve} \quad \exp \sum_{c=0}^0 (\theta - b_{ic}) = 1$$

$b_{ic}$ ; *i*. maddenin *c*. kategorisine ilişkin zorluk parametresidir. Denklem pay kısmında sadece başarıyla tamamlanmış  $k$  tane basamağın zorluk düzeyleri vardır,  $b_{i1}, b_{i2}, \dots, b_{ik}$ . Paydada ise pay kısmındaki ifadenin bir benzeri, olası tüm basamaklar ( $m_i+1$ ) üzerinden hesaplanmaktadır (Wright ve Masters, 1982; Gorin ve ark., 2005). Maddelerin analizi için kullanılacak model belirlendikten sonra, soru bankalarının psikometrik özellikleri; model uyumu, MİF ve tek boyutluluk açısından incelenmiştir.

### Model Uyumunun İncelenmesi

Modele uyumu gösteren istatistikler, gözlenen verinin ölçüm modelinin beklentilerine ne derece uygun olduğunu test eder. Tümel uyum iyiliğinin değerlendirilmesinde üç istatistikten yararlanılır. Bunlardan iki tanesi biri madde diğeri birey için olmak üzere standart normal dağılım gösteren yaklaşık bir z skoruna dönüştürülmüş “madde ve birey etkileşim istatistikleridir”. Bu nedenle maddeler ve bireyler modele uyum gösteriyorsa, yaklaşık olarak 0 ortalama ve 1 standart sapma değerinin elde edilmesi beklenir. Üçüncüsü ki-kare istatistiğine bağlı olarak verilen ve  $\theta$  düzeyi boyunca değişmezlik özelliğini gösteren “madde- $\theta$  etkileşim istatistiğidir”. İstatistiksel olarak anlamlı bir ki-kare değeri, maddelere verilen yanıtların hiyerarşik sıralamasının  $\theta$  düzeyi boyunca değiştiğini gösterir. Bu durum da değişmezlik özelliğinin bozulması anlamına gelir.

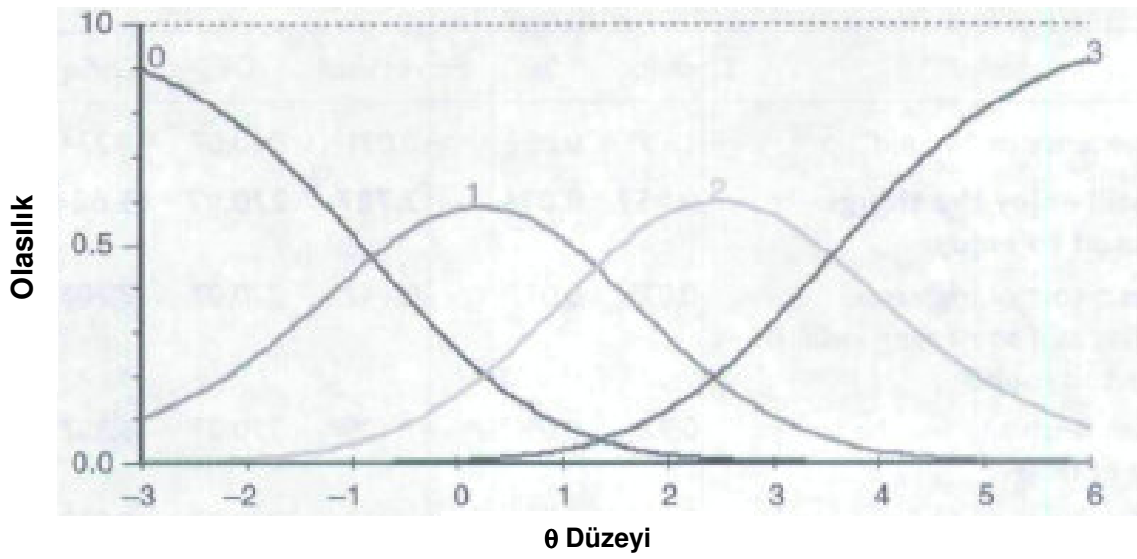
Tümel uyum istatistiklerine ek olarak, her bir birey ve madde için artık değerleri ve ki-kare istatistiği cinsinden hesaplanan uyum istatistikleri bulunmaktadır. Artık değerleri için  $\pm 2.5$  arasında artık değerleri, modele yeterli bir uyum olduğunu gösterir. Ki-kare istatistiğinde ise, modele uyumsuzluk grafiksel olarak gösterilebilir. Bu grafiği çizmek için öncelikle bireyler yetenek düzeylerine göre sınıf aralıkları (Class Intervals) olarak adlandırılan gruplara ayrılırlar. Çalışmadaki sınıf aralığı birey sayısı ile orantılıdır ve her sınıfta yaklaşık 50'şer denek olması istenir. Sınıf aralıkları için gözlenen model uyumu, beklenen model eğrisine (MKF eğrisi) karşı çizilerek uyumsuzluklar belirlenir. Bu çizim sonucunda, iyi uyuma sahip olan maddeler eğrinin üzerinde, aşırı ayırt ediciliğe sahip maddeler eğriden daha dik, düşük ayırt ediciliğe sahip maddeler ise eğriden daha düz bir şekilde sahip olacaktır.

Madde uyumuna ek olarak, birey uyumunun incelenmesi de önemlidir. Uygulama açısından bakıldığında, model beklentilerinden sapmalar gösteren az sayıda birey, madde düzeyinde önemli uyumsuzluklara neden olabilir. Tıp alanında yapılan çalışmalarda, bireyin var olan hastalığının yanı sıra eşlik eden başka hastalıklara da sahip olması bu bireyin modele uyumsuzluk göstermesinde etkili olabilir. Bu nedenle, bazı bireylerin modele uyumsuzluk göstermeleri durumunda, bu durumun nedenlerinin araştırılıp; sorunun çözümlenmesi tercih edilen yaklaşımdır.

Elde edilen soru bankalarının iç tutarlığının bir tahmini olarak birey ayırsama indeksi (BAİ) [*Person Separation Index-(PSI)*] ve Cronbach'ın  $\alpha$  katsayısı hesaplanır. BAİ, uyum istatistiklerine ne derece güvenilebileceğinin de bir göstergesidir. Eğer indeks değeri düşük ise, uyum istatistiklerine güvenilmez. İstatistiksel olarak ayırsama yapılmak istenen grup sayısına bağlı olarak bu değer kabul edilebilir değerleri değişkenlik gösterir. İki grup için 0.70 değeri kabul edilebilirken; 4 grup için bu değer 0.90'dır. Cronbach  $\alpha$  katsayısı, veri setinde eksik gözlem olmadığı durumda elde edilebilirken; 0.70 değeri kabul edilebilir en küçük değerdir.

Soru bankası özelliklerinin geliştirilip geliştirilemeyeceğini görmek için model beklentilerinden sapma kaynakları incelenir. İyi uyuma sahip olan bir modelde, her bir madde için  $\theta$  düzeyi yüksek olan bireylerin yüksek skorlu yanıtları,  $\theta$  düzeyi düşük olan bireylerin de düşük skorlu yanıtları seçmesi beklenir. Rasch analizinde her bir madde için ardışık iki yanıt kategorisinin eşik değerlerinin sıralı olması gerekir. Bir maddeye verilen yanıtları, o maddenin kategori olasılık eğrileri ile incelenebilir. Kısmi kredi modeline iyi uyum gösteren bir madde için,  $\theta$  düzeyi boyunca, her bir yanıt kategorisinin sistematik olarak sırayla en yüksek yanıt olasılığına sahip olması beklenir. Madde uyumsuzluğunun en yaygın kaynaklarından birisi, bireylerin bu yanıt kategorilerini tutarsız olarak seçmeleridir. Bu durum eşik değerlerinin sıra takip etmemesine (*disordered thresholds*) neden olur. Sıra takip etmeyen

eşik değerleri, bireyler yanıt kategorileri arasında tutarlı bir şekilde ayırmama yapmakta güçlük çektikleri durumda ortaya çıkar. Bu problem, gereğinden fazla sayıda kategori olduğunda, kategorilerin adlandırılmasının karışık ya da yanlış anlaşılmaya neden olduğu durumlarda (nadiren, bazen, sıklıkla terimlerinin kullanımı gibi) ortaya çıkabilir. Bu tür durumlarla karşılaşıldığında genellikle, kategorilerin birleştirilmesi ile problem ortadan kaldırılır. Şekil 2.1, eşik değerlerinin sıralı olduğu, Şekil 2.2, eşik değerlerinin sıra takip etmediği örnek birer maddeyi göstermektedir.



Şekil 2.1: Eşik Değerlerinin Sıralı Olduğu Örnek Bir Madde



Şekil 2.2: Eşik Değerlerinin Sıra Takip Etmediği Örnek Bir Madde

### MİF'in İncelenmesi

Model uyumunu etkileyebilecek diğeri bir durum, MİF'dir. MİF, örneklemdaki farklı grupların (örneğin kadın ve erkekler) eşit  $\theta$  düzeyine sahip olmalarına rağmen, belirli bir maddeye farklı şekilde yanıt vermeleri durumunda ortaya çıkar. Örneğin, eşit özürllülük düzeyine sahip olan kadın ve erkekler özürllülük ölçeğinde bir maddeye sistematik olarak farklı biçimde yanıt verebilirler. İki tür MİF tanımlanabilir. Gruplar, bir maddeye verilen yanıtlar bakımından tutarlı sistematik bir farklılık gösteriyorsa "tekdüze (uniform) MİF", göstermiyorsa "tekdüze olmayan (non uniform) MİF"den bahsedilir. Diğeri bir ifadeyle, ölçme aracındaki herhangi bir madde için bir grup,  $\theta$  düzeyi boyunca diğeriinden anlamlı derecede daha yüksek/düşük skora sahip ise tekdüze MİF olduğu söylenir. Diğeri yandan, ölçme aracındaki herhangi bir madde için  $\theta$ 'nın belli bir değerine kadar bir grup, o değerden sonra da diğeri grup anlamlı derecede daha yüksek/düşük skorlara sahip ise tekdüze olmayan MİF'den bahsedilir. Tekdüze MİF olduğunda, her bir grup için ayrı madde elde edilir ve bu maddeler ayrı ayrı kalibre edilir. Tekdüze olmayan MİF için genellikle sorunlu madde ölçme aracından çıkarılır.

MİF varlığı, hem istatistiksel hem de grafiksel olarak belirlenebilir. Bireylere ait değişkenin (örneğin cinsiyet) her bir düzeyi ve  $\theta$ 'nın farklı düzeyleri (sınıf aralıkları) üzerinden her bir madde için varyans analizi gerçekleştirilir. Tekdüze MİF, bireye ilişkin değişken (cinsiyet) için anlamlı ana etki, tekdüze olmayan MİF ise anlamlı etkileşim terimi (cinsiyet x sınıf aralığı) ile belirlenebilir.

### Tek Boyutluluğun İncelenmesi

Modele uyum, eşik değerlerinin sıra takip etmemesi ve MİF sorunları çözümlendiğinde, tek boyutluluğun belirlenmesi için artıklara ilişkin TBA uygulanır. TBA sonuçlarına göre, artıklarda herhangi anlamlı bir yapı

(pattern) olmaması, yerel bağımsızlığı ve sonuç olarak ölçme aracının tek boyutluluk varsayımını sağladığını gösterecektir. Bu varsayım, ilk artık üzerindeki pozitif ve negatif faktör yüklü maddelerden elde edilen tahminlerin bağımsız t testi ile karşılaştırılmasıdır. Eğer, bireylere ilişkin tahmin, pozitif ve negatif gruplarda farklılık gösteriyorsa yerel bağımsızlık varsayımının bozulduğu söylenebilir (Pallant ve Tennant, 2007).

Bu aşamanın sonunda, modele uyum göstermeyen, cinsiyet ve/veya yaş grubu bakımından MIF gösteren ve tek boyutluluk varsayımını bozan maddeler soru bankalarından çıkarılmıştır.

### **Aşama III: Geçerliliğin Belirlenmesi**

Bu aşamada, her bir hastalık grubu için ikinci aşamada elde edilen final soru bankalarının geçerliliği değerlendirilmiştir. Bunun için biri benzetim sonucunda elde edilen veriler, diğeri gerçek veriler üzerinde olmak üzere iki farklı çalışma gerçekleştirilmiştir. Bu çalışmalarda BUT yönteminin uygulanabilmesi için bir program (CAT) yazılmıştır. Bu program temelde gerçek hastalar üzerinde BUT uygulaması yapabilecek şekilde tasarlanmış olup; benzetim çalışmaları için de bir modüle (CATSIM) sahiptir. Bu modül, türetilen veri setleri için otomatik olarak BUT yöntemini uygulamaktadır. BUT yönteminin uygulanması için yazılan bu programda; program analizi ve kod çalışması için Oracle veritabanı kurulmuş; uygulamaların kodları **Oracle 10G Developer Suite** araçları kullanılarak hazırlanmıştır. Bu programda, BUT yönteminin uygulanmasında aşağıdaki temel özellikler benimsenmiştir:

**İlk maddenin seçimi:** BUT yönteminde kullanılacak ilk maddenin seçiminde hastalara ilişkin önsel bir bilgi olmadığı için, ortanca zorluğa sahip olan madde yaklaşımı kullanılmıştır. Bunun için soru bankasında yer alan maddelerden ortanca zorluk değerine en yakın b'ye sahip olan madde başlangıç maddesi olarak hastaya uygulanmıştır.

- **Sonraki maddenin seçimi:** Hastaların özürülük düzeylerinin tahmin edilmesinde en çok olabilirlik yaklaşımı kullanılmıştır. Bu yaklaşımda bir tahmin değerinin elde edilebilmesi için birbirinden farklı en az iki yanıt kategorisinin seçilmesi gerekmektedir. Bu nedenle en çok olabilirlik tahmini elde edilene kadar sıradaki maddenin seçiminde, hastaya uygulanan bir önceki maddenin türü ve bu maddeye verilen yanıt göz önüne alınmıştır:

İki sonuçlu madde için işleyiş: Eğer uygulanan madde iki sonuçlu ise ve hasta yüksek skorlu olan yanıt seçmişse sonraki madde, uygulanan maddenin b'sinden 0.20 daha yüksek olan değere en yakın b'ye sahip olan maddedir. Eğer düşük skorlu olan yanıt seçilmişse sonraki madde, uygulanan maddenin b'sinden 0.20 daha düşük olan değere en yakın b'ye sahip olan maddedir.

Çok sonuçlu madde için işleyiş: Çok sonuçlu bir maddede işleyişi açıklamak için seçilen maddenin kategorileri (0, 1, 2, 3) ve bu kategorilere ilişkin eşik değerleri ( $b_1$ ,  $b_2$ ,  $b_3$ ) biçiminde olsun. Eğer uygulanan madde için hasta 0 kategorisini seçmişse sonraki madde, uygulanan maddenin  $b_1$  değerinden 0.20 daha düşük olan değere en yakın b'ye sahip olan maddedir. Eğer hasta 3 kategorisini seçmişse, sonraki madde uygulanan maddenin  $b_3$  değerine en yakın b'ye sahip olan maddedir. Seçilen kategori 1 veya 2 ise, bu durumda sonraki madde sırasıyla  $b_1$  ve  $b_2$  değerlerinin ortalamasına veya  $b_2$  ve  $b_3$  değerlerinin ortalamasına en yakın b'ye sahip olan maddedir.

- **Hastaların özürülük düzeylerinin tahmin edilmesi:** Hastaların özürülük düzeylerinin tahmini, BUT uygulamalarındaki en önemli aşamalardan birisidir. Bu durum, en son elde edilen  $\theta$  tahminini olduğu kadar, madde seçimi ve testin durdurulmasını da etkiler. Bu çalışmada, hastaların özürülük düzeylerinin tahmin edilmesinde en çok olabilirlik yaklaşımı ve Newton-Raphson algoritması kullanılmıştır. Denklem 2.1'de verilen model kullanıldığında L aşağıdaki gibi elde edilir:

$$L = \prod_{i=1}^n \prod_{k=0}^{m_i} P_{ik}^{u_{ik}} \quad (\text{Denklem 2.2})$$

Bu denklemde  $i$  madde ( $i=1,2,\dots,n$ ),  $k$  kategori ( $k=0,\dots,m_i$ ) indisi ve  $m$  kategori sayısının bir eksiği olmak üzere  $P_{ik}$ ,  $i$ . maddenin  $k$ . kategorisini seçme olasılığıdır.  $u_{ik}$ ,  $i$ . maddenin  $k$ . kategorisi için yanıt yapısını göstermektedir. Kısmi kredi modeline göre  $u_{ik}$ ,  $i$ . maddede seçilen kategori için 1, seçilmeyen kategoriler için 0 değerini alır. Örneğin 1. maddenin yanıt kategorileri 0, 1, 2 ve bireyin bu maddeye verdiği yanıt 2 olsun. Bu durumda  $u_{10}$ ,  $u_{11}$  0 değerini alırken;  $u_{12}$  1 değerini alır.

- **Durdurma kuralı:** BUT uygulamasının sona ermesinde “değişken uzunluk” yaklaşımı kullanılmıştır. Buna göre  $SE(\theta)$  değerleri bakımından ardışık iki uygulama arasındaki farklılık 0.01’in altında ise test sonlandırılmıştır. Hastaların uygulama süresince aynı yanıt kategorilerini seçmeleri durumunda ise test, soru bankasındaki tüm maddeler yanıtlandığında sona ermiştir. Bu durumdaki hastaların özürlülük düzeyleri, yanıt yapılarına bağlı olarak (000 veya 111 gibi) uygulama yapılan gruptaki en düşük/en yüksek  $\theta$  değerine eşit kabul edilmiştir.

### Benzetim Çalışması

Bu amaçla Rasch analizi sonucunda elde edilen madde parametreleri kullanılarak, her bir hastalık grubu için 10000’er hastanın final soru bankalarındaki maddelere verdikleri yanıtlar, RUMMss benzetim programı yardımıyla türetilmiştir. Ayrıca, RUMMss programı ile türetilen bu yanıt yapıları kullanılarak hastaların özürlülük düzeyleri ( $\theta_{\text{Rasch}}$ ) de elde edilebilmiştir.



RUMMss programından türetilen yanıtlarla CATSIM modülü kullanılarak, aynı 10000'er hastanın özürlülük düzeyi ( $\theta_{BUT}$ ) BUT yöntemiyle de elde edilmiştir. Hastaların özürlülük düzeyleri, RUMMss programında soru bankasındaki maddelerin tümü kullanılarak; CATSIM programında ise BUT yöntemi kullanılarak elde edilmiştir.  $\theta_{Rasch}$  ile  $\theta_{BUT}$  değerleri Blant-Altman yöntemi, Spearman ilişki katsayısı ve sınıfiçi korelasyon katsayısı (SKK) [*Intraclass Correlation Coefficient-ICC*] ile karşılaştırılmıştır.

### Gerçek Uygulama

Bu aşamada her bir hastalık grubu için elde edilen final soru bankalarında yer alan maddeler, bel ağrılı 42 ve diz osteoartritli 43'er hastaya klasik kağıt-kalem testi ve BUT yöntemi ile uygulanmıştır. Klasik yöntemin ardından hastalar BUT uygulamasına alınmışlardır. BUT uygulaması öncesinde bilgisayar tecrübelerinin olup olmadığı ve eğitim durumları sorgulanmıştır. Buna göre bilgisayar tecrübesi olan okuryazar hastalar, BUT uygulamasını kendi başlarına tamamlamışlardır. Bilgisayar tecrübesi olmayan ve/veya okuryazar olmayan hastalar BUT yöntemini Biyoistatistik Anabilim Dalında görev yapan bir araştırma görevlisinin yardımıyla tamamlamışlardır. Klasik yöntem ( $\theta_{Rasch}$ ) ve BUT yöntemi ( $\theta_{BUTg}$ ) sonucunda elde edilen özürlülük düzeyleri, benzetim çalışmasındaki yöntemler kullanılarak karşılaştırılmıştır.

### **3. BULGULAR**

Bu çalışmada elde edilen bulgular, bel ağrılı ve diz osteoartritli hastalar için ayrı ayrı, "AFA ve DFA sonuçları", "Rasch analizi sonuçları", "benzetim çalışması ve gerçek uygulama sonuçları" bazında verilecektir.

#### **3.1. Bel Ağrılı Hastalar İçin Elde Edilen Bulgular**

##### **3.1.1. Demografik Bulgular**

Bu çalışma, Ankara Üniversitesi Tıp Fakültesi FTR Anabilim Dalı kliniğinde bel ağrısı tanısı ile tetkik ve tedavi edilen 266 hasta üzerinde yapılmıştır. 266 hastanın 43'ü (%16.2) erkek, 223'ü (%83.8) kadındır. Hastaların yaş ortalaması  $52.2 \pm 12.5$  olarak bulunmuştur. Ortalama şikayet süreleri 8.24 yıl olup; minimum 0.08, maksimum 40 yıldır.

##### **3.1.2. Başlangıç Soru Bankasının Klasik Yöntemle Uygulanması**

Çalışma kapsamında öncelikli olarak WHODAS-II, Oswestry, Roland Morris ve NHP ölçeklerinde yer alan maddeler kullanılarak oluşturulan başlangıç soru bankası (108 madde), 266 hastaya klasik yöntemle uygulanmıştır. Bu uygulama sonucunda maddelere verilen yanıtlar incelenmiş; çalışan ve/veya cinsel olarak aktif olan hasta sayısı çok az olduğundan bu alanlarla ilgili olarak WHODAS-II ölçeğinden 5, Oswestry ölçeğinden 1 tane olmak üzere, toplam 6 madde başlangıç soru bankasından çıkarılmıştır.

### 3.1.3. Açıklayıcı ve Doğrulayıcı Faktör Analizi Sonuçları

İlk aşamada, 102 maddeden oluşan başlangıç soru bankasının faktör yapısını belirlemek için AFA uygulanmıştır. 102 madde ile gerçekleştirilen AFA sonucunda 3 madde diğer maddelerle yüksek negatif korelasyona ( $<-0.90$ ) sahip olduğu için analizden çıkarılmış, kalan 99 madde ile yeni bir AFA uygulanmıştır. Bu analiz sonucunda, maddelerin iki boyutta toplandığı belirlenmiş, faktör yükleri ve ICF sınıflamaları göz önünde bulundurulduğunda ilk boyut "Ağrı-Uyku-Duygusal Durum", ikinci boyut "Aktivite ve Katılım" olarak adlandırılmıştır. "Ağrı-Uyku-Duygusal Durum" boyutunda yer alan maddelerin faktör yükleri 0.425 ile 0.883 arasında, "Aktivite-Katılım" boyutunda yer alan maddelerinki 0.413 ile 0.935 arasında değişkenlik göstermektedir. Bu aşamada hiçbir boyuta yüklenmeyen (her bir boyuttaki faktör yükü  $<0.40$  ve iki faktöre ilişkin yükler arasındaki fark  $<0.10$  olan) 5 madde soru bankasından çıkarılmıştır.

İki faktörlü model için AFA sonucuna göre YHKOK değeri 0.087 olup; bu değer yeterli uyum için kabul edilebilir sınır olan 0.08 değerine oldukça yakındır. AFA sonucunda elde edilen faktörlerin gerçekten bu maddelerden oluşup oluşmadığına karar vermek için DFA yapıldığında KUI değeri 0.869, TLİ değeri 0.918 olarak bulunmuştur. Bu değerlerin 0.90'dan büyük olması kabul edilebilir bir uyum olduğunu gösterdiğinden, elde edilen sonuçların tatmin edici olduğu söylenebilir.

Bu sonuçlara göre "Ağrı-Uyku-Duygusal Durum" boyutundaki madde sayısı 40, "Aktivite-Katılım" boyutundaki madde sayısı 54'tür. AFA ve DFA sonucunda elde edilen soru bankalarındaki maddeler, kodları ve ICF sınıflamaları Çizelge 3.1 ve 3.2'de verilmiştir.

Çizelge 3.1. Bel Ağrılı Hastalar İçin “Ağrı-Uyku-Duygusal Durum” Soru Bankası

Kodu*	Maddeler	ICF Kodu**
wd1.1	Son 30 gün içinde on dakika süreyle bir işe dikkatinizi verebilmede ne kadar güçlük çektiniz?	BF
wd1.2	Son 30 gün içinde yapılması gereken önemli şeyleri akılda tutmada ne kadar güçlük çektiniz?	BF
wd1.3	Son 30 gün içinde günlük yaşamda ortaya çıkan sorunların nedenlerini anlamada ve çözümler bulmada ne kadar güçlük çektiniz?	BF&AP
wd1.4	Son 30 gün içinde yeni işler öğrenmede (yeni bir yol vb.) ne kadar güçlük çektiniz?	BF&AP
wd1.5	Son 30 gün içinde insanların söylediklerini genel olarak anlamada ne kadar güçlük çektiniz?	BF&AP
wd1.6	Son 30 gün içinde bir sohbeti başlatabilmede veya sürdürebilmede ne kadar güçlük çektiniz?	BF&AP
wd4.1	Son 30 gün içinde tanımadığınız insanlarla ilişki kurmada ne kadar güçlük çektiniz?	AP
wd4.2	Son 30 gün içinde bir arkadaşlığı sürdürmede ne kadar güçlük çektiniz?	AP
wd4.3	Son 30 gün içinde yakın olduğunuz insanlarla ilişki kurmada ne kadar güçlük çektiniz?	AP
wd4.4	Son 30 gün içinde yeni arkadaşlar edinmede ne kadar güçlük çektiniz?	AP
wd6.3	Son 30 gün içinde diğer insanların davranış ve tutumları karşısında kendinize saygınızı yitirmeden yaşayabilmekte ne kadar sorunla karşılaştınız?	E
wd6.5	Son 30 gün içinde sağlık durumunuz yüzünden duygusal olarak ne kadar etkilendiniz?	BF
o1	Ağrı şiddeti	BF
o7	Uyku	BF
rm13	Belim hemen hemen her zaman ağrıyor.	BF
rm18	Bel ağrım yüzünden rahat uyuyamıyorum.	BF
rm22	Bel ağrım yüzünden eskisine göre huzursuz ve sinirliyim.	BF
n1	Kendimi sürekli yorgun hissediyorum.	BF
n2	Geceleri ağrım oluyor.	BF
n3	Her şey moralimi bozuyor.	BF
n4	Dayanılmaz şiddette ağrım var.	BF
n5	Uyuyabilmek için ilaç alıyorum.	BF&E
n6	Artık eğlenmeyi unuttum.	AP
n7	Kendimi çok sinirli hissediyorum.	BF
n9	Kendimi yalnız hissediyorum.	BF
n13	Sabahları çok erken saatte uyanıyorum.	BF
n15	İnsanlarla ilişki kurmakta zorlanıyorum.	AP
n16	Günler geçmek bilmiyormuş gibi geliyor.	BF
n20	Bugünlerde çok kolay öfkeleniveriyorum.	BF
n21	Bana yakın hiç kimse yokmuş gibi hissediyorum.	BF
n22	Geceleri çoğunlukla uyanık oluyorum.	BF
n23	Bazen kontrolümü kaybediyormuş gibi hissediyorum.	NC
n28	Sürekli ağrım oluyor.	BF
n29	Uykuya dalabilmek için uzun süre bekliyorum.	BF
n30	Çevremdeki insanlara yük oluyormuşum gibi geliyor.	NC
n31	Geceleri endişelerim yüzünden uyuyamıyorum.	BF
n32	Hayat yaşamaya değmez gibi geliyor.	NC
n33	Gece uykularım çok kötü.	BF
n34	İnsanlarla geçinmek bana zor geliyor.	AP
n37	Sabahları moralim bozuk ve keyifsiz uyanıyorum.	BF

\* wd: WHODAS-II, o: Oswestry, rm: Roland Morris, n: NHP

\*\* BF: Vücut fonksiyonları, AP: Aktivite ve katılım, E: Çevre, NC: ICF tarafından sınıflandırılmayan

Çizelge 3.2. Bel Ağrılı Hastalar İçin “Aktivite-Katılım” Soru Bankası

Kodu*	Maddeler	ICF Kodu**
wd2.1	Son 30 gün içinde 30 dakikada gibi uzun süre ayakta durmada ne kadar güçlük çektiniz?	AP
wd2.2	Son 30 gün içinde otururken ayağa kalkmada ne kadar güçlük çektiniz?	AP
wd2.3	Son 30 gün içinde evin içinde dolaşmada ne kadar güçlük çektiniz?	AP
wd2.4	Son 30 gün içinde evden dışarı çıkmada ne kadar güçlük çektiniz?	AP
wd2.5	Son 30 gün içinde bir kilometre dolayında uzun mesafe yürümede ne kadar güçlük çektiniz?	AP
wd3.1	Son 30 gün içinde banyo yapmada ne kadar güçlük çektiniz?	AP
wd3.2	Son 30 gün içinde giyinmede ne kadar güçlük çektiniz?	AP
wd3.3	Son 30 gün içinde beslenmede ne kadar güçlük çektiniz?	AP
wd3.4	Son 30 gün içinde birkaç gün yalnız kalmada ne kadar güçlük çektiniz?	AP
wd5.2	Son 30 gün içinde evle ilgili sorumlulukları yerine getirmede ne kadar güçlük çektiniz?	AP
wd5.3	Son 30 gün içinde evdeki önemli görevlerinizi iyi bir şekilde yapmada ne kadar güçlük çektiniz?	AP
wd5.4	Son 30 gün içinde yapılması gereken tüm ev işlerini bitirmede ne kadar güçlük çektiniz?	AP
wd5.5	Son 30 gün içinde ev işlerini yeterince hızlı yapmada ne kadar güçlük çektiniz?	AP
wd6.1	Son 30 gün içinde sosyal faaliyetlere herkes kadar katılabilmekte (şenlikler, eğlenceler, dini faaliyetler vb.) ne kadar güçlük çektiniz?	AP
wd6.2	Son 30 gün içinde çevrenizden kaynaklanan engeller yüzünden ne kadar sorunla karşılaştınız?	E
wd6.8	Son 30 gün içinde dinlenmek veya eğlenmek amacıyla kendi başınıza bir şeyler yapmakta ne kadar güçlük çekiyorsunuz?	AP
o2	Kişisel bakım (yıkama, giyinme vs.)	BF&AP
o3	Ağırlık kaldırma	BF&AP
o4	Yürüme	BF&AP&E
o5	Oturma	BF&AP
o6	Ayakta durma	BF&AP
o9	Sosyal yaşam	BF&AP
o10	Gezi	BF&AP
rm1	Bel ağrım yüzünden zamanımın büyük çoğunluğunu evde geçiriyorum.	BF&AP
rm2	Belimi rahatlatmak için sık sık ayakta duruş, oturuş veya yatış şeklimi değiştirmek zorunda kalıyorum.	BF&AP
rm3	Bel ağrım yüzünden eskisinden daha yavaş yürüyorum.	BF&AP
rm4	Bel ağrım yüzünden evde yaptığım birçok işi artık yapmıyorum.	BF&AP
rm5	Bel ağrım yüzünden merdivenleri çıkarken tirabzanlara tutunuyorum.	BF&AP&E
rm6	Bel ağrım yüzünden dinlenmek için sık sık uzanıyorum.	BF (AP)
rm7	Bel ağrım yüzünden sandalyeden kalkarken bir yere tutunmak ihtiyacı duyuyorum.	BF&E AP
rm8	Bel ağrım yüzünden bazı işlerimi başkalarına yaptırıyorum.	BF&E (AP)
rm9	Bel ağrım yüzünden eskisinden daha yavaş giyiniyorum.	BF&AP
rm10	Bel ağrım yüzünden sadece kısa süre ayakta kalabiliyorum.	BF&AP
rm11	Bel ağrım yüzünden eğilmekten ve çömelmekten kaçınıyorum.	BF&AP
rm12	Bel ağrım yüzünden sandalyeden kalkarken zorluk çekiyorum.	BF (AP)
rm14	Bel ağrım yüzünden yatakta dönmekte zorluk çekiyorum.	BF&AP
rm16	Bel ağrım yüzünden çoraplarımı giymekte zorluk çekiyorum.	BF&AP
rm17	Bel ağrım yüzünden sadece kısa mesafeleri yürüyebiliyorum.	BF&AP
rm19	Bel ağrım yüzünden bir başkasının yardımıyla giyiniyorum.	BF&AP&E
rm21	Bel ağrım yüzünden evdeki ağır işleri yapmaktan kaçınıyorum.	BF&AP
rm23	Bel ağrım yüzünden merdivenleri her zamankinden daha yavaş çıkıyorum.	BF&AP

\* wd: WHODAS-II, o: Oswestry, rm: Roland Morris, n: NHP

\*\* BF: Vücut fonksiyonları, AP: Aktivite ve katılım, E: Çevre, NC: ICF tarafından sınıflandırılmayan

Çizelge 3.2. (Devam) Bel Ağrılı Hastalar İçin “Aktivite-Katılım” Soru Bankası

Kodu*	Maddeler	ICF Kodu**
rm24	Bel ağrım yüzünden zamanın çoğunu yatakta geçiriyorum.	BF&AP
n8	Hareket etmek, pozisyon değiştirmek bana ağrı veriyor.	BF&AP
n11	Öne eğilmek benim için zor oluyor.	AP
n12	En basit işler için bile caba göstermem gerekiyor.	BF
n17	Merdiven inip çıkmakta zorlanıyorum.	AP
n18	Bazı şeylere, yerlere uzanmak yetişmek zor oluyor.	AP
n19	Yürürken ağrım oluyor.	AP
n24	Ayakta durunca ağrım oluyor.	BF&AP
n25	Kendi kendime giyinmek zor oluyor.	AP
n26	Çabucak yoruluyor.	BF
n27	Uzun süre ayakta durmak bana zor geliyor (örneğin mutfakta veya otobüs beklerken gibi)	AP
n36	Merdiven inip çıkarken ağrım oluyor.	BF&AP
n38	Otururken ağrım oluyor.	BF&AP

\* wd: WHODAS-II, o: Oswestry, rm: Roland Morris, n: NHP

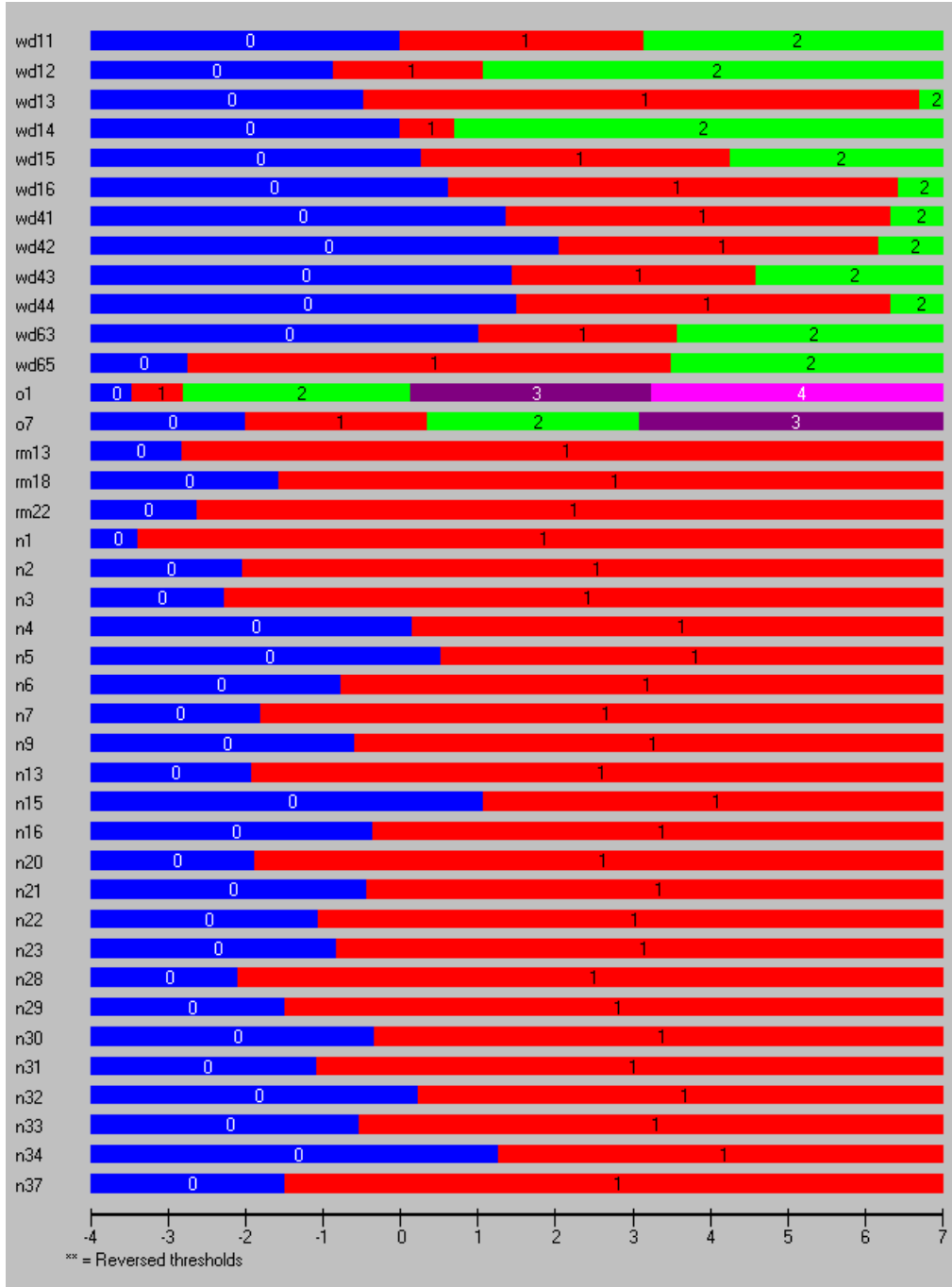
\*\* BF: Vücut fonksiyonları, AP: Aktivite ve katılım, E: Çevre, NC: ICF tarafından sınıflandırılmayan

### 3.1.4. Rasch Analizi Sonuçları

Bu aşamada, AFA ve DFA sonucunda elde edilen soru bankalarının psikometrik özellikleri kısmi kredi modeli kullanılarak incelenmiştir. Buna göre, Rasch analizi sonucunda modele uyum göstermeyen, cinsiyet ve/veya yaş grubu ( $\leq 46$ , 47-52, 53-58,  $\geq 59$ ) bakımından MİF gösteren ve tek boyutluluk varsayımını bozan maddeler çalışmadan çıkarılmış; BUT yöntemi için kullanılacak olan final soru bankaları elde edilmiştir. Bu bölümde, Rasch analizi sonuçları herbir soru bankası için ayrı ayrı ele alınacaktır.

#### 3.1.4.1. “Ağrı-Uyku-Duygusal Durum” Soru Bankasının Rasch Analizi

“Ağrı-Uyku-Duygusal Durum” soru bankasının Rasch analizinde ilk olarak, maddelere verilen yanıtlar, kategori olasılık eğrileri ile incelenmiş, eşik değerleri sıralı olmayan maddelerin bazı yanıt kategorileri birleştirilmiştir. Soru bankasındaki 40 madde için eşik değerlerine ilişkin birleştirme sonrasındaki yapı Şekil 3.1’de verilmiştir.



Şekil 3.1. “Ağrı-Uyku-Duygusal Durum” Soru Bankasındaki Maddelerin Eşik Değerleri

Şekil 3.1 incelendiğinde soru bankasındaki maddelerin yanıt kategori sayıları eşit olmamakla beraber, eşik değerleri arasındaki uzaklık da birbirine eşit değildir. Bu durum da Rasch analizinde neden kısmi kredi modelinin kullanıldığını daha açık ifade etmektedir.

İkinci aşamada, soru bankasında yer alan 40 maddenin model uyumları incelenmiştir. Bunlar içerisinde wd12, wd14, o1, o7, n3, n5 ve n33 kodlu maddelerin ki-kare değerlerine ilişkin Bonferroni düzeltmeli olasılık (P) değerleri incelendiğinde modele uyumsuzluk gösterdikleri belirlenmiş, bu nedenle de soru bankasından çıkarılmışlardır. Hastalara ilişkin artık değerleri incelendiğinde modele uyum göstermeyen hasta bulunmamıştır. Bu aşamada 33 maddeden oluşan soru bankasının tek boyutlu olup olmadığı artıklar üzerinden yapılan TBA yardımıyla incelenmiştir. İlk temel bileşen üzerinde pozitif ve negatif yüklene maddelerin ayrı birer boyut oluşturup oluşturmadığı test edildiğinde tek boyutluluk varsayımını bozan herhangi bir yapı gözlenmemiştir. Yerel bağımsızlık varsayımının sağlanıp sağlanmadığını test etmek için artık korelasyon matrisi incelenmiş, 0.50'nin üzerinde korelasyona sahip olan sadece iki madde olduğu için bu varsayımın sağlandığı belirlenmiştir. Cinsiyet ve yaş grubu bakımından MİF olup olmadığı değerlendirildiğinde maddelerin sorunlu olmadığı görülmüştür.

Bu sonuçlara göre elde edilen 33 maddelik “Ağrı-Uyku-Duygusal Durum” final soru bankasının tümel uyum iyiliği istatistikleri incelendiğinde madde etkileşim istatistiği ortalaması (standart sapması) 0.552 (0.992), birey etkileşim istatistiği ortalaması (standart sapması) -0.379 (1.077)'dir. Bu değerlerin ortalaması 0, standart sapması 1 değerine yakın olduğu için soru bankasında yer alan maddelerin ve hastaların modele yeterli uyum gösterdiği söylenir. “Madde- $\theta$  etkileşim istatistiği” incelendiğinde ki-kare değeri (P değeri) 132.638 (0.013587) olduğundan ve Bonferroni düzeltmesine göre anlamlı olmadığından bu soru bankası için maddelere verilen yanıtların hiyerarşik sıralamasının  $\theta$  düzeyi boyunca değişmediği yani değişmezlik özelliğinin sağlandığı sonucuna ulaşılır.

Tümel uyum istatistiklerine ek olarak, “Ağrı-Uyku-Duygusal Durum” final soru bankasında yer alan her bir madde için artık değerleri ve ki-kare istatistiği cinsinden hesaplanan uyum istatistikleri Çizelge 3.3'te sunulmuştur.



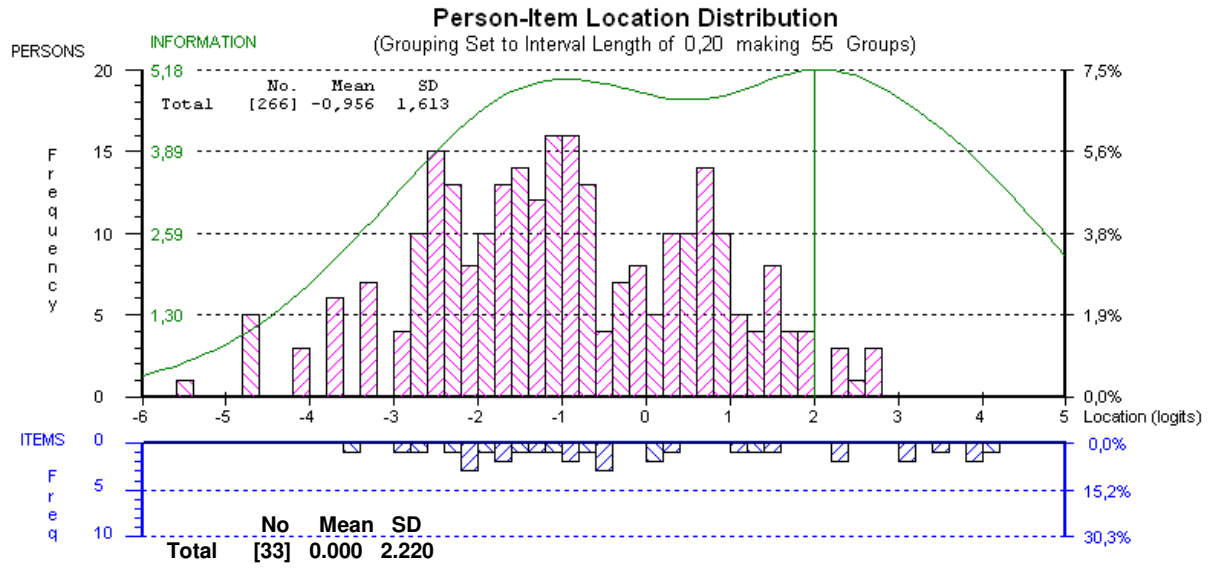
Çizelge 3.3. “Ağrı-Uyku-Duygusal Durum” Final Soru Bankasındaki Maddelerin Uyum İstatistikleri

Maddeler	b	Standart Hata	Artık	Serbestlik Derecesi	Ki-Kare	Serbestlik Derecesi	Olasılık
wd1.1	1.518	0.140	0.879	255.610	8.326	3	0.040
wd1.3	3.084	0.151	0.042	255.610	0.498	3	0.919
wd1.5	2.245	0.153	-0.067	255.610	1.549	3	0.671
wd1.6	3.521	0.166	-0.147	255.610	0.985	3	0.805
wd4.1	3.878	0.196	0.383	240.180	3.269	3	0.352
wd4.2	4.110	0.221	-0.495	253.690	7.708	3	0.052
wd4.3	3.084	0.185	-1.006	255.610	1.652	3	0.648
wd4.4	3.953	0.208	0.021	226.680	2.977	3	0.395
wd6.3	2.264	0.162	0.725	253.690	1.592	3	0.661
wd6.5	0.317	0.159	-0.237	252.720	2.529	3	0.470
rm13	-2.956	0.179	-0.266	254.650	3.331	3	0.343
rm18	-1.661	0.152	0.655	254.650	1.915	3	0.590
rm22	-2.762	0.173	-1.237	254.650	1.950	3	0.583
n1	-3.530	0.203	-0.454	254.650	1.845	3	0.605
n2	-2.146	0.159	0.671	254.650	3.030	3	0.387
n4	0.093	0.159	1.423	254.650	3.156	3	0.368
n6	-0.859	0.150	-0.427	254.650	3.987	3	0.263
n7	-1.948	0.156	-2.143	254.650	8.312	3	0.040
n9	-0.680	0.151	-1.264	254.650	6.513	3	0.089
n13	-2.018	0.157	1.478	254.650	11.751	3	0.008
n15	1.062	0.180	-1.854	254.650	8.311	3	0.040
n16	-0.427	0.153	-2.530	254.650	9.019	3	0.029
n20	-2.018	0.157	-0.733	254.650	0.632	3	0.889
n21	-0.512	0.152	-1.411	253.690	8.458	3	0.037
n22	-1.129	0.150	-1.092	254.650	5.465	3	0.141
n23	-0.904	0.150	-1.229	254.650	0.889	3	0.828
n28	-2.208	0.160	-0.693	254.650	3.152	3	0.369
n29	-1.590	0.152	-0.154	254.650	6.140	3	0.105
n30	-0.420	0.153	-0.630	254.650	0.055	3	0.997
n31	-1.200	0.150	-2.118	254.650	5.686	3	0.128
n32	0.189	0.160	-0.895	254.650	1.724	3	0.632
n34	1.249	0.186	-0.187	254.650	3.736	3	0.291
n37	-1.601	0.152	-1.579	254.650	2.496	3	0.476

Çizelge 3.3’e göre artık değerleri  $\pm 2.5$  arasında değerler aldığından ve ki-kare değerlerine ilişkin olasılık değerleri Bonferroni düzeltmeli P değerinden ( $0.05/33=0.0015$ ) daha yüksek olduğundan soru bankasındaki tüm maddelerin modele uyum gösterdiği söylenir.

Soru bankasının iç tutarlılığının incelenmesinde, Cronbach  $\alpha$  katsayısı ve BAİ kullanılmaktadır. Ancak veri setinde kayıp gözlem olduğu için, soru bankasının iç tutarlılığının incelenmesinde sadece BAİ kullanılmış ve bu indeks 0.912 olarak elde edilmiştir. Bu değer yorumlandığında elde edilen soru bankasının güvenilir olduğu sonucuna ulaşılır.

“Ağrı-Uyku-Duygusal Durum” final soru bankasında yer alan maddelerin ve hastaların  $\theta$  düzeyi boyunca dağılımı Şekil 3.2’de verilmiştir.



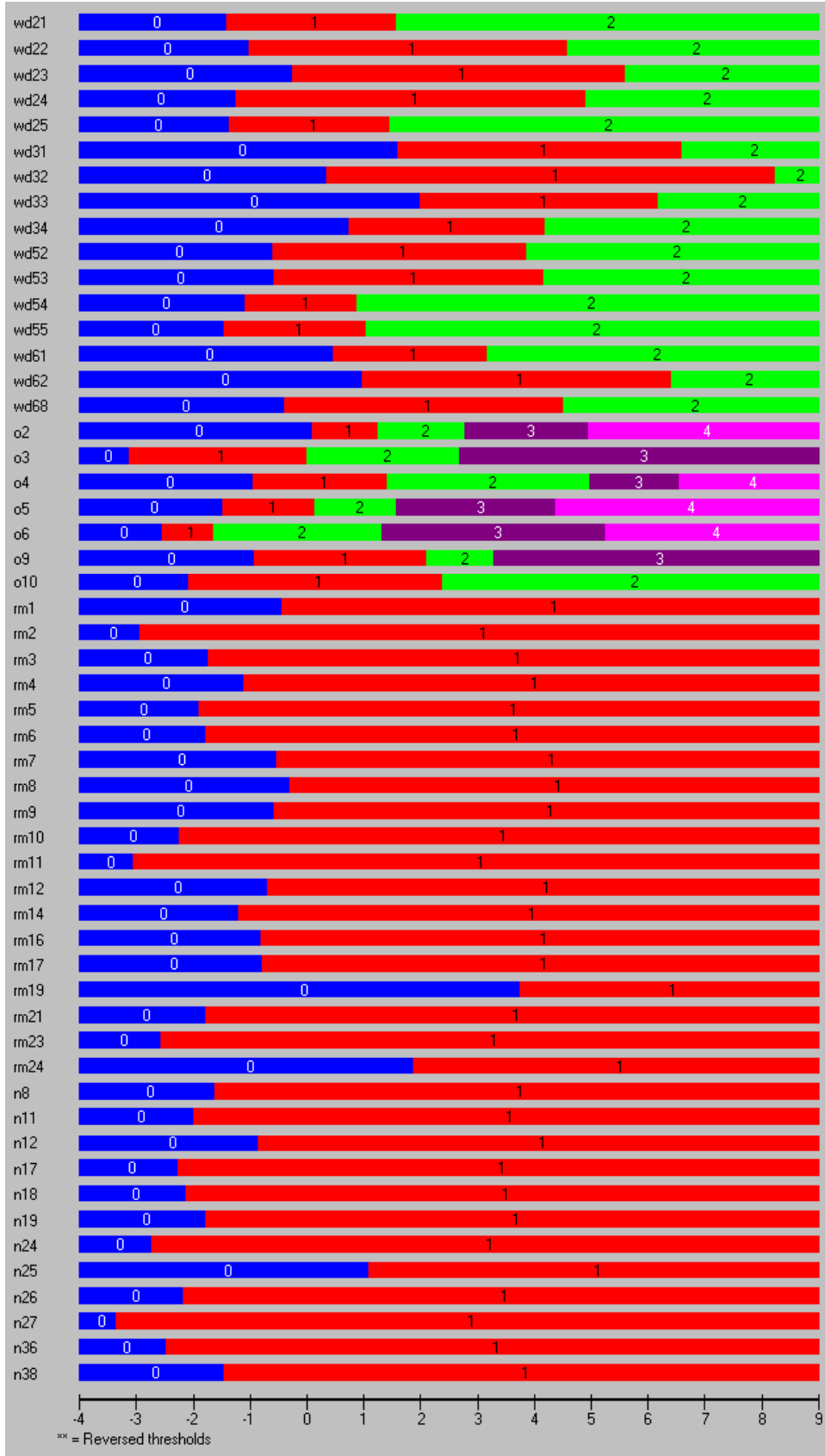
Şekil 3.2. “Ağrı-Uyku-Duygusal Durum” Final Soru Bankasındaki Maddelerin ve Hastaların  $\theta$  Düzeyi Boyunca Dağılımı

Soru bankasında yer alan maddelerin b değerlerinin ortalaması ile incelenen grubun ortalamasının birbirine uygun olması, soru bankasının  $\theta$ 'nın tüm düzeylerinde bilgi verici olması hedeflenir. Şekil 3.2'nin önemli bir özelliği, maddelerin ve hastaların görece konumlarının belirlenebilmesidir. Buna göre, maddelerin hastaların konumları tarafından hemen hemen kapsandığı, fakat bu durumun tersinin doğru olmadığı yani “Ağrı-Uyku-Duygusal Durum” final soru bankasının  $\theta$ 'nın tüm düzeylerinde yeterli sayıda madde içermediği sonucuna ulaşılır. Bu yüzden soru bankasının,  $\theta$  düzeyi -3.5 ile 4 arasında olan hastalar için daha fazla bilgi sağladığı söylenir.

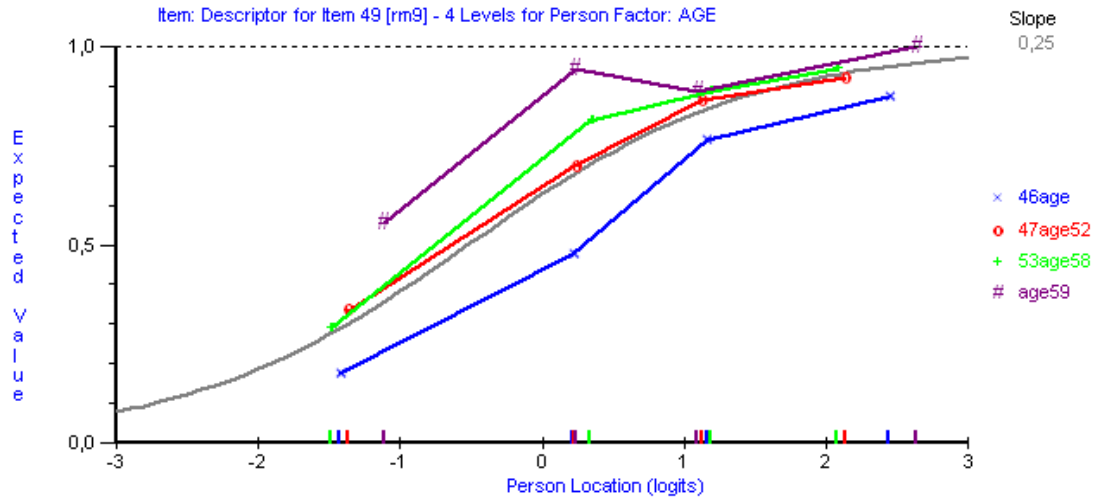
### 3.1.4.2. “Aktivite-Katılım” Soru Bankasının Rasch Analizi

“Aktivite-Katılım” soru bankasının Rasch analizinde ilk olarak, maddelere verilen yanıtlar, kategori olasılık eğrileri ile incelenmiş, eşik değerleri sıralı olmayan maddelerin bazı yanıt kategorileri birleştirilmiştir. Soru bankasındaki 54 madde için eşik değerlerine ilişkin birleştirme sonrasındaki yapı Şekil 3.3’te verilmiştir. Şekil 3.3 incelendiğinde soru bankasındaki maddelerin yanıt kategori sayılarının yanı sıra, eşik değerleri arasındaki uzaklığın da birbirine eşit olmadığı görülmektedir.

İkinci aşamada, soru bankasında yer alan 54 maddenin model uyumları incelenmiştir. Bunlar içerisinde o2, o3 ve o5 kodlu maddelerin ki-kare değerlerine ilişkin Bonferroni düzeltmeli olasılık (P) değerleri incelendiğinde modele uyumsuzluk gösterdikleri belirlenmiş, bu nedenle de soru bankasından çıkarılmışlardır. Hastalara ilişkin artık değerleri incelendiğinde modele uyum göstermeyen hasta bulunmamıştır. Bu aşamada 51 maddeden oluşan soru bankasının tek boyutlu olup olmadığı artıklar üzerinden yapılan TBA yardımıyla incelenmiştir. İlk temel bileşen üzerinde pozitif ve negatif yüklenen maddelerin ayrı birer boyut oluşturup oluşturmadığı test edildiğinde tek boyutluluk varsayımını bozan wd52 ve n17 kodlu maddeler analizden çıkarılmıştır. Yerel bağımsızlık varsayımının sağlanıp sağlanmadığını test etmek için artık korelasyon matrisi incelenmiş, 0.50’nin üzerinde korelasyona sahip olan sadece iki madde olduğu için bu varsayımın sağlandığı belirlenmiştir. Cinsiyet ve yaş grubu bakımından MİF olup olmadığı grafiksel ve istatistiksel olarak değerlendirildiğinde cinsiyete göre hiçbir madde MİF göstermezken; rm9 kodlu maddenin yaş grubuna göre tekdüze MİF gösterdiği belirlenmiştir. Bu maddenin yaş gruplarına göre MKF eğrileri Şekil 3.4’te gösterilmiştir.



Şekil 3.3. “Aktivite-Katılım” Soru Bankasındaki Maddelerin Eşik Değerleri



Şekil 3.4. “rm9” Kodlu Maddenin Yaş Gruplarına Göre MKF Eğrileri

Şekil 3.4 incelendiğinde, rm9 kodlu “Bel ağrım yüzünden eskisinden daha yavaş giyiniyorum” maddesinin 59 yaşının üzerinde olan hastalar için daha zor bir faaliyet olduğu sonucuna ulaşılır. Bu madde, hastanın kendine bakımı ile ilgili özürülük düzeyini sorguladığından soru bankasında kalmasının gerekli olduğu düşünülmüştür. Ayrıca MİF varlığının denek sayısı ile bağlantılı olabileceği düşünülerek, denek sayısı artırılarak bu maddenin MİF gösterip göstermediğinin tekrar incelenmesine karar verilmiştir.

Bu sonuçlara göre elde edilen 49 maddelik “Aktivite-Katılım” final soru bankasının tümel uyum iyiliği istatistikleri incelendiğinde madde etkileşim istatistiği ortalaması (standart sapması) -0.236 (1.375), birey etkileşim istatistiği ortalaması (standart sapması) -0.397 (0.937)’dir. Bu değerlerin ortalaması 0, standart sapması 1 değerine yakın olduğu için soru bankasında yer alan maddelerin ve hastaların modele yeterli uyum gösterdiği söylenir. Madde- $\theta$  etkileşim istatistiği incelendiğinde ki-kare değeri (P değeri) 190.712 (0.008882) olduğundan ve Bonferroni düzeltmesine göre anlamlı olmadığından bu soru bankası için değişmezlik özelliğinin sağlandığı yorumuna ulaşılır. Tümel uyum istatistiklerine ek olarak, “Aktivite-Katılım” final soru bankasında yer alan her bir madde için artık değerleri ve ki-kare istatistiği cinsinden hesaplanan uyum istatistikleri Çizelge 3.4’te sunulmuştur.

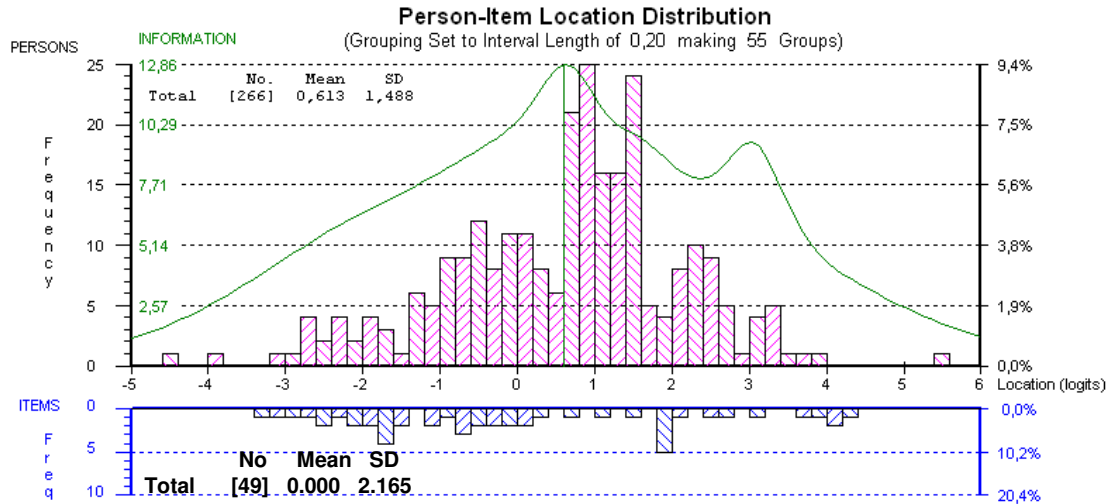
Çizelge 3.4. “Aktivite-Katılım” Final Soru Bankasındaki Maddelerin Uyum İstatistikleri

Maddeler	b	Standart Hata	Artık	Serbestlik Derecesi	Ki-Kare	Serbestlik Derecesi	Olasılık
wd2.1	0.132	0.116	2.141	259.030	6.255	3	0.100
wd2.2	1.846	0.152	-1.500	259.030	2.695	3	0.441
wd2.3	2.752	0.146	-0.579	258.050	3.289	3	0.349
wd2.4	1.907	0.162	-1.600	259.030	5.242	3	0.155
wd2.5	0.110	0.114	-0.115	259.030	1.716	3	0.633
wd3.1	4.182	0.149	-0.844	259.030	2.264	3	0.519
wd3.2	4.376	0.145	-0.638	259.030	1.780	3	0.619
wd3.3	4.163	0.156	1.320	259.030	0.275	3	0.965
wd3.4	2.540	0.130	2.009	254.160	7.187	3	0.066
wd5.3	1.882	0.146	-1.405	232.730	4.658	3	0.199
wd5.4	-0.038	0.115	-0.758	230.790	5.290	3	0.152
wd5.5	-0.144	0.120	-0.608	230.790	2.190	3	0.534
wd6.1	1.911	0.119	0.471	256.100	0.456	3	0.928
wd6.2	3.722	0.142	1.391	257.080	6.266	3	0.099
wd6.8	2.115	0.141	-0.986	256.100	0.793	3	0.851
o4	3.082	0.106	0.361	259.030	7.119	3	0.068
o6	0.651	0.104	0.447	259.030	4.730	3	0.193
o9	1.578	0.102	1.585	258.050	11.801	3	0.008
o10	0.216	0.137	-1.365	248.310	1.964	3	0.580
rm1	-0.396	0.156	-0.836	258.050	2.526	3	0.471
rm2	-2.904	0.274	-0.207	258.050	2.106	3	0.551
rm3	-1.705	0.198	-1.807	258.050	9.433	3	0.024
rm4	-1.064	0.173	-0.307	258.050	2.586	3	0.460
rm5	-1.862	0.205	-1.482	258.050	4.863	3	0.182
rm6	-1.745	0.200	-0.001	258.050	1.760	3	0.624
rm7	-0.482	0.158	0.522	258.050	2.695	3	0.441
rm8	-0.252	0.153	1.712	258.050	8.550	3	0.036
rm9	-0.522	0.159	-0.079	258.050	1.594	3	0.661
rm10	-2.219	0.225	-1.558	258.050	4.238	3	0.237
rm11	-3.023	0.284	-0.823	258.050	2.227	3	0.527
rm12	-0.647	0.162	-0.527	258.050	3.033	3	0.387
rm14	-1.157	0.177	-1.006	257.080	4.851	3	0.183
rm16	-0.753	0.164	-0.316	258.050	2.821	3	0.420
rm17	-0.745	0.164	-2.068	258.050	8.215	3	0.042
rm19	3.879	0.253	-0.611	258.050	4.508	3	0.212
rm21	-1.734	0.199	0.091	258.050	1.217	3	0.749
rm23	-2.556	0.247	-1.294	258.050	3.386	3	0.336
rm24	1.959	0.157	0.786	258.050	7.979	3	0.046
n8	-1.594	0.193	-0.478	258.050	5.599	3	0.133
n11	-1.954	0.210	-1.284	258.050	2.197	3	0.533
n12	-0.805	0.166	-1.470	258.050	7.025	3	0.071
n18	-2.090	0.217	-1.710	258.050	3.055	3	0.383

Çizelge 3.4. (Devam) “Aktivite-Katılım” Final Soru Bankasındaki Maddelerin Uyum İstatistikleri

Maddeler	b	Standart Hata	Artık	Serbestlik Derecesi	Ki-Kare	Serbestlik Derecesi	Olasılık
n19	-1.740	0.200	-1.263	258.050	3.802	3	0.284
n24	-2.701	0.257	-0.611	258.050	0.774	3	0.856
n25	1.174	0.146	1.298	258.050	2.948	3	0.400
n26	-2.134	0.220	5.455	258.050	1.989	3	0.575
n27	-3.334	0.314	-1.092	258.050	1.178	3	0.758
n36	-2.459	0.240	-1.310	258.050	5.191	3	0.158
n38	-1.417	0.186	1.374	258.050	2.394	3	0.495

Çizelge 3.4’e göre artık değerleri  $\pm 2.5$  arasında değerler aldığından ve ki-kare değerlerine ilişkin olasılık değerleri Bonferroni düzeltmeli P değerinden ( $0.05/49=0.0010$ ) daha yüksek olduğu için soru bankasındaki tüm maddelerin modele uyum gösterdiği söylenir. Soru bankasının iç tutarlılığının incelenmesinde, veri setinde kayıp gözlem olduğu için, sadece BAİ kullanılmış ve bu indeks 0.936 olarak elde edilmiştir. Bu değer yorumlandığında elde edilen soru bankasının güvenilir olduğu sonucuna ulaşılır. “Aktivite-Katılım” final soru bankasında yer alan maddelerin ve hastaların  $\theta$  düzeyi boyunca dağılımı Şekil 3.5’te verilmiştir.



Şekil 3.5. “Aktivite-Katılım” Final Soru Bankasındaki Maddelerin ve Hastaların  $\theta$  Düzeyi Boyunca Dağılımı

Soru bankasında yer alan maddelerin b değerlerinin ortalaması ile incelenen grubun ortalamasının birbirine uygun olması, soru bankasının  $\theta$ 'nın tüm düzeylerinde bilgi verici olduğunu gösterir. Buna göre, Şekil 3.5 incelendiğinde hem maddelerin, hastaların konumları tarafından kapsandığı, hem de bu durumun tersinin doğru olduğu yani “Aktivite-Katılım” final soru bankasının  $\theta$ 'nın tüm düzeylerinde yeterli sayıda madde içerdiği sonucuna ulaşılır. Soru bankasının,  $\theta$  düzeyi -3.5 ile 4.5 arasında olan hastalar için bilgi sağladığı söylenir.

### **3.1.5. Benzetim Çalışması ve Gerçek Uygulama Sonuçları**

Benzetim çalışmasında, Rasch analizi sonucunda elde edilen madde parametreleri kullanılarak RUMMss programı ile 10000 hastanın yanıtları türetilmiştir. Aynı zamanda bu yanıt yapıları ile hastaların özürlülük düzeyleri ( $\theta_{\text{Rasch}}$ ) de elde edilmiştir. RUMMss programından türetilen yanıtlarla CATSIM modülü kullanılarak, 10000 hastanın özürlülük düzeyi ( $\theta_{\text{BUT}}$ ) BUT yöntemiyle de bulunmuştur.  $\theta_{\text{Rasch}}$  ile  $\theta_{\text{BUT}}$  değerleri Blant-Altman yöntemi, Spearman ilişki katsayısı ve SKK ile karşılaştırılmıştır. Gerçek uygulamada ise, elde edilen final soru bankalarında yer alan maddeler, 42 hastaya klasik yöntem ve BUT yöntemi ile uygulanmış; her iki uygulama sonucunda hastalara ilişkin elde edilen özürlülük düzeyleri (sırasıyla  $\theta_{\text{Raschg}}$  ve  $\theta_{\text{BUTg}}$ ) benzetim çalışmasında kullanılan yöntemlerle karşılaştırılmıştır. Bu bölümde, elde edilen sonuçlar herbir soru bankası için ayrı ayrı ele alınacaktır.



### 3.1.5.1. “Ağrı-Uyku-Duygusal Durum” Soru Bankası için Benzetim Çalışması ve Gerçek Uygulama Sonuçları

#### Benzetim Çalışması

“Ağrı-Uyku-Duygusal Durum” final soru bankası ile yapılan BUT uygulamasında 10000 hastanın 9044’ü (%90.4) için özürülük düzeyi belirlenebilirken; 108’i (%1.1) için en düşük özürülük düzeyi ataması yapılmış (108 hasta BUT sürecinde tüm maddelere en düşük kategoride yanıt vermişlerdir), 848 (%8.5) hasta için de Newton-Raphson algoritmasından kaynaklanan sorunlar nedeniyle özürülük düzeyi elde edilememiştir. Bu nedenle “Ağrı-Uyku-Duygusal Durum” soru bankasına ilişkin karşılaştırmalar, 9152 (%91.5) hasta üzerinden yapılmıştır.

Elde edilen sonuçlar Blant-Altman yöntemi ile karşılaştırıldığında, “Ağrı-Uyku-Duygusal Durum” final soru bankası için  $\theta_{\text{Rasch}}$  ve  $\theta_{\text{BUT}}$  değerleri arasındaki farkların %95’i (-1.652, 1.161) uyum sınırları içerisindedir. Uyum sınırları incelendiğinde BUT ile elde edilen özürülük düzeylerinin %5’inin, Rasch analizi ile elde edilenlerden 1.652 lojit daha düşük ya da 1.161 lojit daha yüksek bulunabileceği yorumuna ulaşılır. Buna göre 9152 hastanın 8656’sı (%94.6) uyum sınırları içerisinde kalırken; 496 hastanın (%5.4) 227’sinde BUT yöntemi ile elde edilen; 269’unda Rasch analizi ile elde edilen özürülük düzeyi daha yüksek bulunmuştur.

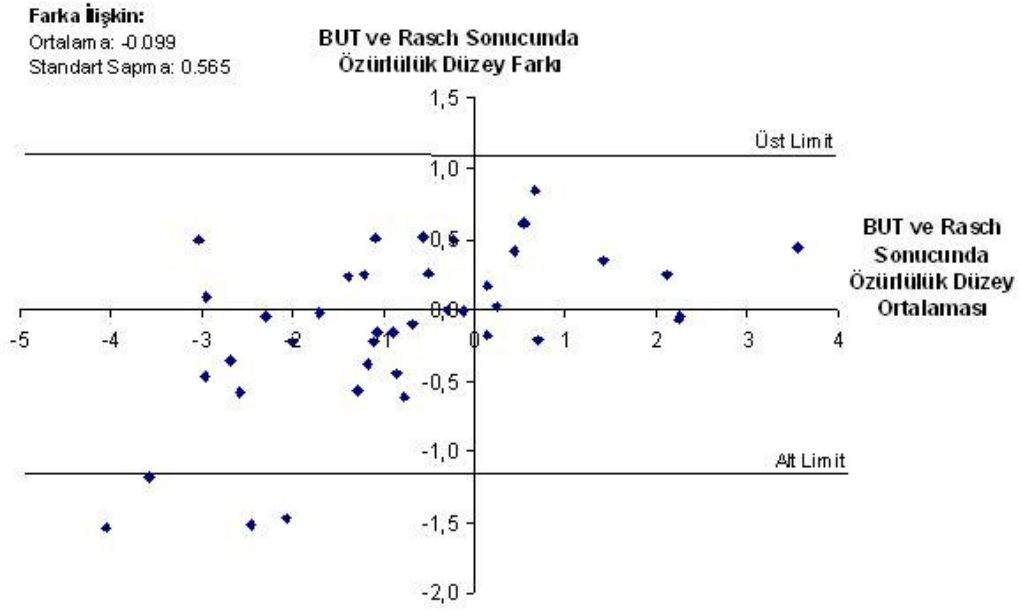
$\theta_{\text{Rasch}}$  ile  $\theta_{\text{BUT}}$  değerleri Spearman ilişki katsayısı ile karşılaştırıldığında, iki değer arasında pozitif yönlü bir ilişki belirlenmiştir ( $r=0.95$ ,  $p<0.001$ ). Aynı değerler için SKK 0.93 ( $p<0.001$ ) olarak bulunmuştur. Bu katsayı için 0.85-0.94 arasındaki değerlerin, yüksek uyumu gösterdiği bilgisi ışığında Rasch analizi ile elde edilen özürülük düzeyleri ile BUT yöntemi ile elde edilenlerin çok uyumlu olduğu sonucuna ulaşılır. Hastaların özürülük düzeyleri,

RUMMss programında soru bankasındaki maddelerin tümü kullanılarak; CATSIM programında ise BUT yöntemi kullanılarak elde edildiğinden yukarıdaki sonuçlar madde sayısı bakımından da değerlendirilmiştir. Buna göre  $\theta_{\text{Rasch}}$  değerleri 33 madde üzerinden elde edilirken;  $\theta_{\text{BUT}}$  değerleri ortalama 17 (standart sapma: 4, minimum-maksimum: 6-33) madde kullanılarak bulunmuştur.

### Gerçek Uygulama

“Ağrı-Uyku-Duygusal Durum” final soru bankası ile gerçek hastalar üzerinde yapılan BUT uygulamasında  $\theta_{\text{Raschg}}$  ve  $\theta_{\text{BUTg}}$  değerleri arasındaki farkların %95’i (-1.207, 1.008) Blant-Altman uyum sınırları içerisindedir. Uyum sınırları incelendiğinde BUT yöntemi ile elde edilen özürülük düzeylerinin %5’inin, Rasch analizi ile elde edilenlerden 1.207 lojit daha düşük ya da 1.008 lojit daha yüksek bulunabileceği yorumuna ulaşılır. Buna göre 42 hastanın 39’u (%93) uyum sınırları içerisinde kalırken; 3 hasta (%7) için Rasch analizi ile elde edilen özürülük düzeyi daha yüksek bulunmuştur. Gerçek uygulama sonuçlarına göre “Ağrı-Uyku-Duygusal Durum” final soru bankası için uyum sınırlarına ilişkin Blant-Altman grafiği Şekil 3.6’da verilmiştir.

$\theta_{\text{Raschg}}$  ile  $\theta_{\text{BUTg}}$  değerleri Spearman ilişki katsayısı ile karşılaştırıldığında, iki değer arasında pozitif yönlü bir ilişki belirlenmiştir ( $r=0.96$ ,  $p<0.001$ ). Aynı değerler için SKK 0.95 ( $p<0.001$ ) olarak bulunmuştur. Bu katsayı için 0.95-1.00 arasındaki değerlerin, mükemmel uyumu gösterdiği bilgisi ışığında Rasch analizi ile elde edilen özürülük düzeyleri ile BUT yöntemi ile elde edilenlerin çok uyumlu olduğu sonucuna ulaşılır.  $\theta_{\text{Raschg}}$  değerleri 33 madde üzerinden elde edilirken;  $\theta_{\text{BUTg}}$  değerleri ortalama 18 (standart sapma: 3, minimum-maksimum: 6-21) madde kullanılarak bulunmuştur.



Şekil 3.6. Gerçek Uygulama Sonuçlarına Göre “Ağrı-Uyku-Duygusal Durum” Final Soru Bankası İçin Uyum Sınırlarına İlişkin Blant-Altman Grafiği

### 3.1.5.2. “Aktivite-Katılım” Soru Bankası için Benzetim Çalışması ve Gerçek Uygulama Sonuçları

#### Benzetim Çalışması

“Aktivite-Katılım” final soru bankası ile yapılan BUT uygulamasında 10000 hastanın 9507’si (%95.1) için özürlülük düzeyi belirlenebilirken; 54’ü (%0.5) için en düşük özürlülük düzeyi ataması yapılmış (54 hasta BUT sürecinde tüm maddelere en düşük kategoride yanıt vermişlerdir), 439 (%4.4) hasta için de Newton-Raphson algoritmasından kaynaklanan sorunlar nedeniyle özürlülük düzeyi elde edilememiştir. Bu nedenle “Aktivite-Katılım” soru bankasına ilişkin karşılaştırmalar, 9561 (%95.6) hasta üzerinden yapılmıştır.

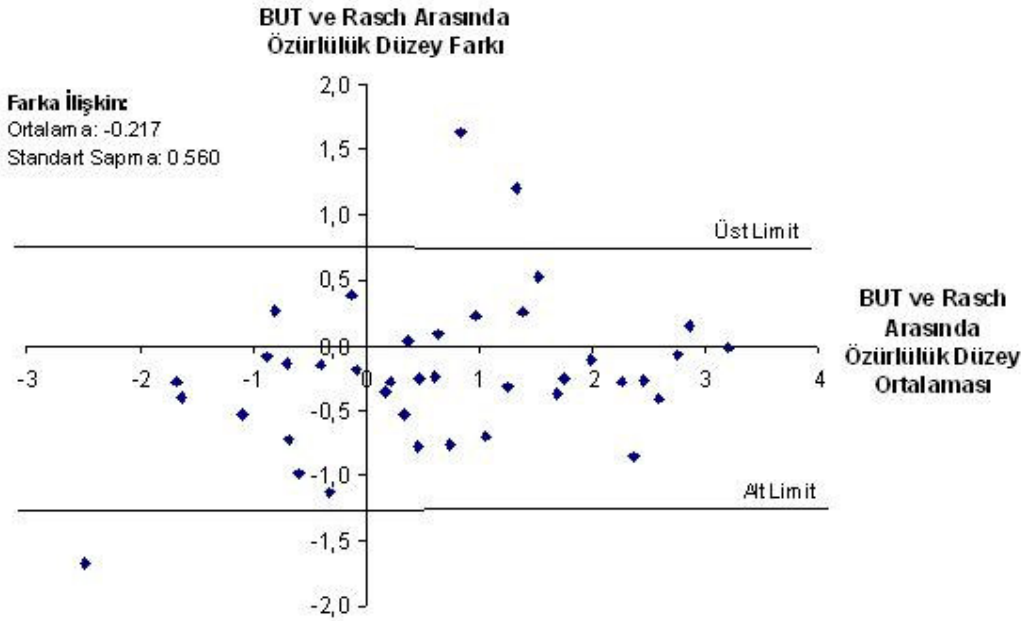
“Aktivite-Katılım” final soru bankası için  $\theta_{\text{Rasch}}$  ve  $\theta_{\text{BUT}}$  değerleri arasındaki farkların %95’i (-1.274, 1.381) Blant-Altman uyum sınırları içerisinde. Uyum sınırları incelendiğinde BUT yöntemi ile elde edilen özürülük düzeylerinin %5’inin, Rasch analizi ile elde edilenlerden 1.274 lojit daha düşük ya da 1.381 lojit daha yüksek bulunabileceği yorumuna ulaşılır. Buna göre 9561 hastanın 9020’si (%94) uyum sınırları içerisinde kalırken; 541 hastanın (%6) 278’inde BUT yöntemi ile elde edilen; 263’ünde Rasch analizi ile elde edilen özürülük düzeyi daha yüksek bulunmuştur.

$\theta_{\text{Rasch}}$  ile  $\theta_{\text{BUT}}$  değerleri Spearman ilişki katsayısı ile karşılaştırıldığında, iki değer arasında pozitif yönlü bir ilişki belirlenmiştir ( $r=0.97$ ,  $p<0.001$ ). Aynı değerler için SKK 0.95 ( $p<0.001$ ) olarak bulunmuştur. Bu katsayı için 0.95-1.00 arasındaki değerlerin, mükemmel uyumu gösterdiği bilgisi ışığında Rasch analizi ile elde edilen özürülük düzeyleri ile BUT yöntemi ile elde edilenlerin çok uyumlu olduğu sonucuna ulaşılır. Hastaların özürülük düzeyleri, RUMMss programında soru bankasındaki maddelerin tümü kullanılarak; CATSIM programında ise BUT yöntemi kullanılarak elde edildiğinden yukarıdaki sonuçlar madde sayısı bakımından da değerlendirilmiştir. Buna göre  $\theta_{\text{Rasch}}$  değerleri 49 madde üzerinden elde edilirken;  $\theta_{\text{BUT}}$  değerleri ortalama 17 (standart sapma: 4, minimum-maksimum: 5-49) madde kullanılarak bulunmuştur.

### Gerçek Uygulama

“Aktivite-Katılım” final soru bankası ile gerçek hastalar üzerinde yapılan BUT uygulamasında  $\theta_{\text{Rasch}}$  ve  $\theta_{\text{BUT}}$  değerleri arasındaki farkların %95’i (-1.313, 0.880) Blant-Altman uyum sınırları içerisinde. Uyum sınırları incelendiğinde BUT yöntemi ile elde edilen özürülük düzeylerinin %5’inin, Rasch analizi ile elde edilenlerden 1.313 lojit daha düşük ya da 0.880 lojit daha yüksek bulunabileceği yorumuna ulaşılır. Buna göre 42 hastanın 39’u (%93) uyum sınırları içerisinde kalırken; 3 hastanın (%7) 2’sinde BUT yöntemi ile elde

edilen, 1'inde Rasch analizi ile elde edilen özürlülük düzeyi daha yüksek bulunmuştur. Gerçek uygulama sonuçlarına göre "Aktivite-Katılım" final soru bankası için uyum sınırlarına ilişkin Blant-Altman grafiği Şekil 3.7'de verilmiştir.



Şekil 3.7. Gerçek Uygulama Sonuçlarına Göre "Aktivite-Katılım" Final Soru Bankası İçin Uyum Sınırlarına İlişkin Blant-Altman Grafiği

$\theta_{Raschg}$  ile  $\theta_{BUTg}$  değerleri Spearman ilişki katsayısı ile karşılaştırıldığında, iki değer arasında pozitif yönlü bir ilişki belirlenmiştir ( $r=0.93$ ,  $p<0.001$ ). Aynı değerler için SKK 0.91 ( $p<0.001$ ) olarak bulunmuştur. Bu katsayı için 0.85-0.94 arasındaki değerlerin, yüksek uyumu gösterdiği bilgisi ışığında Rasch analizi ile elde edilen özürlülük düzeyleri ile BUT yöntemi ile elde edilenlerin çok uyumlu olduğu sonucuna ulaşılır.  $\theta_{Raschg}$  değerleri soru bankasındaki maddelerin tümü kullanılarak; CAT programında ise BUT yöntemi kullanılarak elde edildiğinden yukarıdaki sonuçlar madde sayısı bakımından da değerlendirilmiştir. Buna göre  $\theta_{Raschg}$  değerleri 49 madde üzerinden elde edilirken;  $\theta_{BUTg}$  değerleri ortalama 17 (standart sapma: 2, minimum-maksimum: 10-20) madde kullanılarak bulunmuştur.

## **3.2. Diz Osteoartritli Hastalar İçin Elde Edilen Bulgular**

### **3.2.1. Demografik Bulgular**

Bu çalışma, Ankara Üniversitesi Tıp Fakültesi FTR Anabilim Dalı kliniğinde diz osteoartriti tanısı ile tetkik ve tedavi edilen 225 hasta üzerinde yapılmıştır. 225 hastanın 49'u (%21.8) erkek, 176'sı (%78.2) kadındır. Hastaların yaş ortalaması  $58.0 \pm 10.9$  olarak bulunmuştur. Ortalama şikayet süreleri 6.04 yıl olup; minimum 0.08, maksimum 40 yıldır.

### **3.2.2. Başlangıç Soru Bankasının Klasik Yöntemle Uygulanması**

Çalışma kapsamında öncelikli olarak WHODAS-II, WOMAC, HAQ ve NHP ölçeklerinde yer alan maddeler kullanılarak oluşturulan başlangıç soru bankası (118 madde), 225 hastaya klasik yöntemle uygulanmıştır. Bu uygulama sonucunda maddelere verilen yanıtlar incelenmiş;

- Çalışan
- Cinsel olarak aktif olan
- Küvette banyo yapan
- Ağır ev işi yapan

hasta sayısı çok az olduğundan, bu alanlarla ilgili olarak WHODAS-II ölçeğinden 5, WOMAC ölçeğinden 2 ve HAQ ölçeğinden 1 tane olmak üzere toplam 8 madde başlangıç soru bankasından çıkarılmıştır.

### 3.2.3. Açıklayıcı ve Doğrulayıcı Faktör Analizi Sonuçları

İlk aşamada, 110 maddeden oluşan başlangıç soru bankasının faktör yapısını belirlemek için AFA uygulanmıştır. 110 madde ile gerçekleştirilen AFA sonucunda maddelerin iki boyutta toplandığı belirlenmiş, faktör yükleri ve ICF sınıflamaları göz önünde bulundurulduğunda ilk boyut “Aktivite-Ağrı”, ikinci boyut “Ağrı-Sosyal-Bilişsel-Duygusal Durum” olarak adlandırılmıştır. “Aktivite-Ağrı” boyutunda yer alan maddelerin faktör yükleri 0.401 ile 0.894 arasında, “Ağrı-Sosyal-Bilişsel-Duygusal Durum” boyutunda yer alan maddelerinki 0.405 ile 0.877 arasında değişkenlik göstermektedir. Bu aşamada hiçbir boyuta yüklenmeyen (her bir boyuttaki faktör yükü  $<0.40$  ve iki faktöre ilişkin yükler arasındaki fark  $<0.10$  olan) 4 madde soru bankasından çıkarılmıştır. Her iki boyuttaki faktör yükü  $>0.40$  olan 5 madde, her iki boyutta da yer almıştır.

İki faktörlü model için AFA sonucuna göre YHKOK değeri 0.111 olup; bu değer yeterli uyum için kabul edilebilir sınır olan 0.08 değere çok yakın olmasa da yeterli kabul edilmiştir. AFA sonucunda elde edilen faktörlerin gerçekten bu maddelerden oluşup oluşmadığına karar vermek için DFA yapıldığında, KUI değeri 0.843, TLI değeri 0.933 olarak bulunmuştur. Bu değerlerin 0.90’dan büyük olması kabul edilebilir bir uyum olduğunu gösterdiğinden, elde edilen sonuçların tatmin edici olduğu söylenebilir.

Bu sonuçlara göre “Aktivite-Ağrı” boyutundaki madde sayısı 73, “Ağrı-Sosyal-Bilişsel-Duygusal Durum” boyutundaki madde sayısı 38’dir. AFA ve DFA sonucunda elde edilen soru bankalarında yer alan maddeler, kodları ve ICF sınıflamaları Çizelge 3.5 ve 3.6’da verilmiştir.

Çizelge 3.5. Diz Osteoartritli Hastalar İçin “Aktivite-Ağrı” Soru Bankası

Kodu*	Maddeler	ICF Kodu**
v1	Son 2 gün içinde düz bir zeminde yürürken ne kadar ağrınız olduğunu belirtiniz.	BF
v2	Son 2 gün içinde merdiven çıkarken ve inerken ne kadar ağrınız olduğunu belirtiniz.	BF
v3	Son 2 gün içinde gece yatağınızda iken uykunuzu bozan ağrının ne kadar olduğunu belirtiniz.	BF
v4	Son 2 gün içinde oturur veya yatar haldeyken ne kadar ağrınız olduğunu belirtiniz.	BF
v5	Son 2 gün içinde ayakta dururken ne kadar ağrınız olduğunu belirtiniz.	BF
v6	Son 2 gün içinde sabah uyandıktan hemen sonra hissettiğiniz tutukluğun şiddetini belirtiniz.	BF
v7	Son 2 gün içinde günün ilerleyen saatlerinde oturduktan, yattıktan veya dinlendikten sonra hissettiğiniz tutukluğun şiddetini belirtiniz.	BF&AP
v8	Merdiven inerken ne kadar güçlük çekiyorsunuz?	AP
v9	Merdiven çıkarken ne kadar güçlük çekiyorsunuz?	AP
v10	Oturduğunuz yerden kalkarken ne kadar güçlük çekiyorsunuz?	AP
v11	Ayakta dururken ne kadar güçlük çekiyorsunuz?	AP
v12	Yere eğilirken ne kadar güçlük çekiyorsunuz?	AP
v13	Düz bir zeminde yürürken ne kadar güçlük çekiyorsunuz?	AP
v14	Arabaya veya otobüse binip inerken ne kadar güçlük çekiyorsunuz?	AP
v15	Alışveriş yaparken ne kadar güçlük çekiyorsunuz?	AP
v16	Çoraplarınızı/dizaltı çoraplarınızı/külotlu çorabınızı giyerken ne kadar güçlük çekiyorsunuz?	AP
v17	Son 2 gün içinde yataktan kalkarken ne kadar güçlük çektiniz?	AP
v18	Son 2 gün içinde çoraplarınızı/dizaltı çoraplarınızı/külotlu çorabınızı çıkartırken ne kadar güçlük çektiniz?	AP
v19	Son 2 gün içinde yatakta yatarken ne kadar güçlük çektiniz?	AP
v21	Son 2 gün içinde otururken ne kadar güçlük çektiniz?	AP
v22	Son 2 gün içinde tuvalete oturup kalkarken ne kadar güçlük çektiniz?	AP
v24	Son 2 gün içinde hafif ev işleri yaparken ne kadar güçlük çektiniz?	AP
n2	Geceleri ağrım oluyor.	BF
n4	Dayanılmaz şiddette ağrım var.	BF
n8	Hareket etmek, pozisyon değiştirmek bana ağrı veriyor.	BF&AP
n10	Sadece ev içinde yürüyebiliyorum.	AP
n11	Öne eğilmek benim için zor oluyor.	AP
n14	Hiç yürüyemiyorum	AP
n17	Merdiven inip çıkmakta zorlanıyorum.	AP
n18	Bazı şeylere, yerlere uzanmak yetişmek zor oluyor.	AP
n19	Yürürken ağrım oluyor.	AP
n24	Ayakta durunca ağrım oluyor.	BF&AP
n25	Kendi kendime giyinmek zor oluyor.	AP
n27	Uzun süre ayakta durmak bana zor geliyor (örneğin mutfakta veya otobüs beklerken gibi)	AP
n28	Sürekli ağrım oluyor.	BF
n36	Merdiven inip çıkarken ağrım oluyor.	BF&AP
n38	Otururken ağrım oluyor.	BF&AP
wd2.1	Son 30 gün içinde 30 dakikada gibi uzun süre ayakta durmada ne kadar güçlük çektiniz?	AP
wd2.2	Son 30 gün içinde otururken ayağa kalkmada ne kadar güçlük çektiniz?	AP

\* v: WOMAC, n: NHP, wd: WHODAS-II, h: HAQ

\*\* BF: Vücut fonksiyonları, AP: Aktivite ve katılım, E: Çevre, NC: ICF tarafından sınıflandırılmayan



Çizelge 3.5. (Devam) Diz Osteoartritli Hastalar İçin “Aktivite-Ağrı” Soru Bankası

Kodu*	Maddeler	ICF Kodu**
wd2.3	Son 30 gün içinde evin içinde dolaşmada ne kadar güçlük çektiniz?	AP
wd2.4	Son 30 gün içinde evden dışarı çıkmada ne kadar güçlük çektiniz?	AP
wd2.5	Son 30 gün içinde bir kilometre dolayında uzun mesafe yürümede ne kadar güçlük çektiniz?	AP
wd3.1	Son 30 gün içinde banyo yapmada ne kadar güçlük çektiniz?	AP
wd3.2	Son 30 gün içinde giyinmede ne kadar güçlük çektiniz?	AP
wd3.3	Son 30 gün içinde beslenmede ne kadar güçlük çektiniz?	AP
wd3.4	Son 30 gün içinde birkaç gün yalnız kalmada ne kadar güçlük çektiniz?	AP
wd5.2	Son 30 gün içinde evle ilgili sorumlulukları yerine getirmede ne kadar güçlük çektiniz?	AP
wd5.3	Son 30 gün içinde evdeki önemli görevlerinizi iyi bir şekilde yapmada ne kadar güçlük çektiniz?	AP
wd5.4	Son 30 gün içinde yapılması gereken tüm ev işlerini bitirmede ne kadar güçlük çektiniz?	AP
wd5.5	Son 30 gün içinde ev işlerini yeterince hızlı yapmada ne kadar güçlük çektiniz?	AP
wd6.1	Son 30 gün içinde sosyal faaliyetlere herkes kadar katılabilmekte (şenlikler, eğlenceler, dini faaliyetler vb.) ne kadar güçlük çektiniz?	AP
wd6.2	Son 30 gün içinde çevrenizden kaynaklanan engeller yüzünden ne kadar sorunla karşılaştınız?	E
wd6.5	Son 30 gün içinde sağlık durumunuz yüzünden duygusal olarak ne kadar etkilendiniz?	BF
wd6.8	Son 30 gün içinde dinlenmek veya eğlenmek amacıyla kendi başınıza bir şeyler yapmakta ne kadar güçlük çekiyorsunuz?	AP
h1	Ayakkabı bağlamak ve düğme iliklemek dahil, kendi kendinize giyinebiliyor musunuz?	AP
h2	Saçınızı yıkayabiliyor musunuz?	AP
h3	Düz bir sandalyeden kalkabiliyor musunuz?	AP
h4	Yatağa yatıp, kalkabiliyor musunuz?	AP
h5	Etinizi kesebiliyor musunuz?	AP
h6	Dolu bir fincanı veya bardağı ağızınıza götürebiliyor musunuz?	AP
h7	Yeni bir karton süt veya meyve suyu kutusunu açabiliyor musunuz?	AP
h8	Dışarıda, düz bir zemin üzerinde yürüyebiliyor musunuz?	AP
h9	Beş basamak merdiven çıkabiliyor musunuz?	AP
h10	Kendi kendinize yıkanıp, kurulanabiliyor musunuz?	AP
h12	Tuvalete oturup kalkabiliyor musunuz?	AP
h13	Başınızın biraz üzerinde duran 2,5 kilo ağırlığındaki bir nesneye (örneğin şeker torbası gibi) uzanıp, nesneyi aşağıya indirebiliyor musunuz?	AP
h14	Eğilip yerden bir giysiyi alabiliyor musunuz?	AP
h15	Araba kapılarını açabiliyor musunuz?	AP
h16	Daha önceden açılmış olan kavanoz kapaklarını açabiliyor musunuz?	AP
h17	Muslukları açıp kapatabiliyor musunuz?	AP
h18	Günlük işlere koşturup, alışveriş yapabiliyor musunuz?	AP
h19	Arabaya binip inebiliyor musunuz?	AP
h20	Yerleri süpürme veya bahçe işleri gibi günlük işleri yapabiliyor musunuz?	AP

\* v: WOMAC, n: NHP, wd: WHODAS-II, h: HAQ

\*\* BF: Vücut fonksiyonları, AP: Aktivite ve katılım, E: Çevre, NC: ICF tarafından sınıflandırılmayan

Çizelge 3.6. Diz Osteoartritli Hastalar İçin “Ağrı-Sosyal-Bilişsel-Duygusal Durum” Soru Bankası

Kodu*	Maddeler	ICF Kodu**
n1	Kendimi sürekli yorgun hissediyorum.	BF
n2	Geceleri ağrım oluyor.	BF
n3	Her şey moralimi bozuyor.	BF
n4	Dayanılmaz şiddette ağrım var.	BF
n5	Uyuyabilmek için ilaç alıyorum.	BF&E
n6	Artık eğlenmeyi unuttum.	AP
n7	Kendimi çok sinirli hissediyorum.	BF
n8	Hareket etmek, pozisyon değiştirmek bana ağrı veriyor.	BF&AP
n9	Kendimi yalnız hissediyorum.	BF
n12	En basit işler için bile çaba göstermem gerekiyor.	BF
n13	Sabahları çok erken saatte uyanıyorum.	BF
n15	İnsanlarla ilişki kurmakta zorlanıyorum.	BF
n16	Günler geçmek bilmiyormuş gibi geliyor.	BF
n20	Bugünlerde çok kolay öfkeleniveriyorum.	BF
n21	Bana yakın hiç kimse yokmuş gibi hissediyorum.	BF
n22	Geceleri çoğunlukla uyanık oluyorum.	BF
n23	Bazen kontrolümü kaybediyormuş gibi hissediyorum.	NC
n26	Çabucak yoruluyorum.	BF
n28	Sürekli ağrım oluyor.	BF
n29	Uykuya dalabilmek için uzun süre bekliyorum.	BF
n30	Çevremdeki insanlara yük oluyormuşum gibi geliyor.	NC
n31	Geceleri endişelerim yüzünden uyuyamıyorum.	BF
n32	Hayat yaşamaya değmez gibi geliyor.	NC
n33	Gece uykularım çok kötü.	BF
n34	İnsanlarla geçinmek bana zor geliyor.	AP
n37	Sabahları moralim bozuk ve keyifsiz uyanıyorum.	BF
n38	Otururken ağrım oluyor.	BF&AP
wd1.1	Son 30 gün içinde on dakika süreyle bir işe dikkatinizi verebilmede ne kadar güçlük çektiniz?	BF
wd1.2	Son 30 gün içinde yapılması gereken önemli şeyleri akılda tutmada ne kadar güçlük çektiniz?	BF
wd1.3	Son 30 gün içinde günlük yaşamda ortaya çıkan sorunların nedenlerini anlamada ve çözümler bulmada ne kadar güçlük çektiniz?	BF&AP
wd1.4	Son 30 gün içinde yeni işler öğrenmede (yeni bir yol vb.) ne kadar güçlük çektiniz?	BF&AP
wd1.5	Son 30 gün içinde insanların söylediklerini genel olarak anlamada ne kadar güçlük çektiniz?	BF&AP
wd1.6	Son 30 gün içinde bir sohbeti başlatabilmede veya sürdürebilmede ne kadar güçlük çektiniz?	BF&AP
wd4.1	Son 30 gün içinde tanımadığınız insanlarla ilişki kurmada ne kadar güçlük çektiniz?	AP
wd4.2	Son 30 gün içinde bir arkadaşlığı sürdürmede ne kadar güçlük çektiniz?	AP
wd4.3	Son 30 gün içinde yakın olduğunuz insanlarla ilişki kurmada ne kadar güçlük çektiniz?	AP
wd4.4	Son 30 gün içinde yeni arkadaşlar edinmede ne kadar güçlük çektiniz?	AP
wd6.3	Son 30 gün içinde diğer insanların davranış ve tutumları karşısında kendinize saygınızı yitirmeden yaşayabilmekte ne kadar sorunla karşılaştınız?	E

\* v: WOMAC, n: NHP, wd: WHODAS-II, h: HAQ

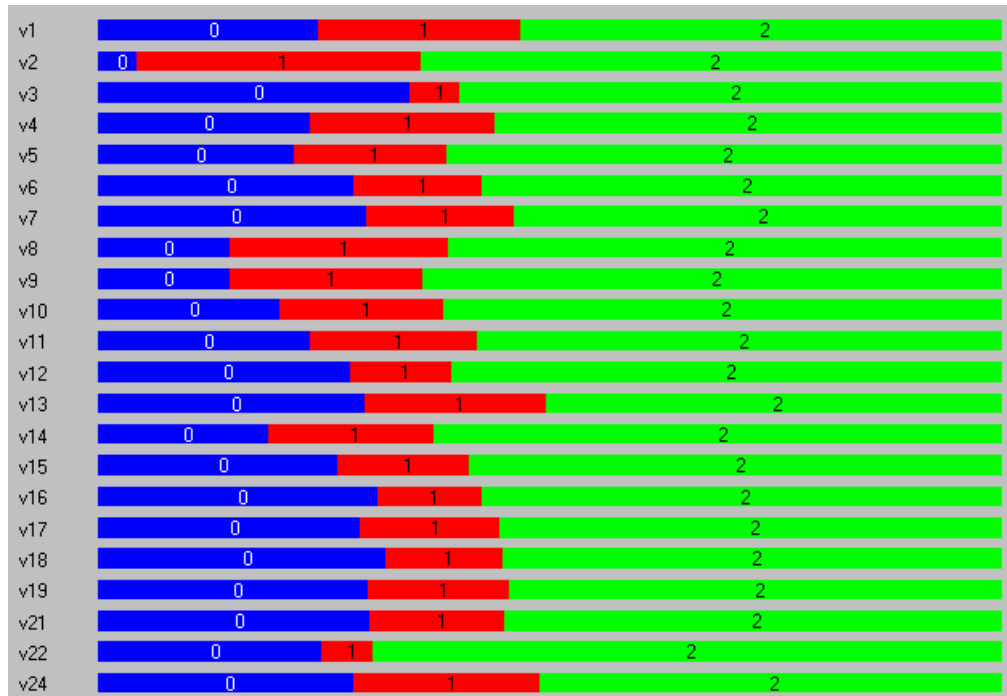
\*\* BF: Vücut fonksiyonları, AP: Aktivite ve katılım, E: Çevre, NC: ICF tarafından sınıflandırılmayan

### 3.2.4. Rasch Analizi Sonuçları

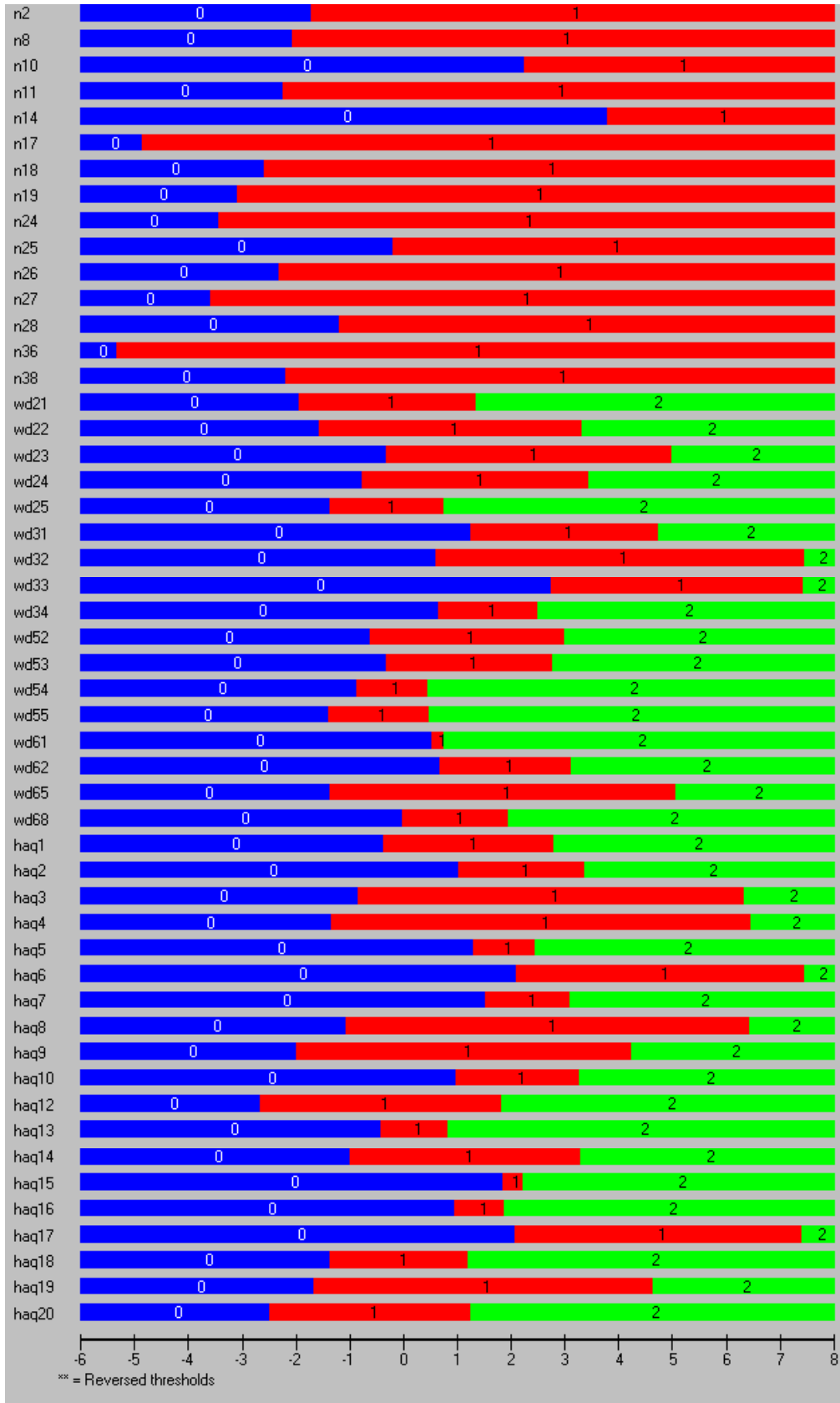
Bu aşamada, AFA ve DFA sonucunda elde edilen soru bankalarının psikometrik özellikleri kısmi kredi modeli kullanılarak incelenmiştir. Buna göre, Rasch analizi sonucunda modele uyum göstermeyen, cinsiyet ve/veya yaş grubu ( $\leq 53$ , 54-63,  $\geq 64$ ) bakımından MİF gösteren ve tek boyutluluk varsayımını bozan maddeler çalışmadan çıkarılmış; BUT yöntemi için kullanılacak olan final soru bankaları elde edilmiştir. Bu bölümde, Rasch analizi sonuçları herbir soru bankası için ayrı ayrı ele alınacaktır.

#### 3.2.4.1. “Aktivite-Ağrı” Soru Bankasının Rasch Analizi

“Aktivite-Ağrı” soru bankasının Rasch analizinde ilk olarak, maddelere verilen yanıtlar, kategori olasılık eğrileri ile incelenmiş, eşik değerleri sıralı olmayan maddelerin bazı yanıt kategorileri birleştirilmiştir. Soru bankasındaki 73 madde için eşik değerlerine ilişkin birleştirme sonrasındaki yapı Şekil 3.8’de verilmiştir.



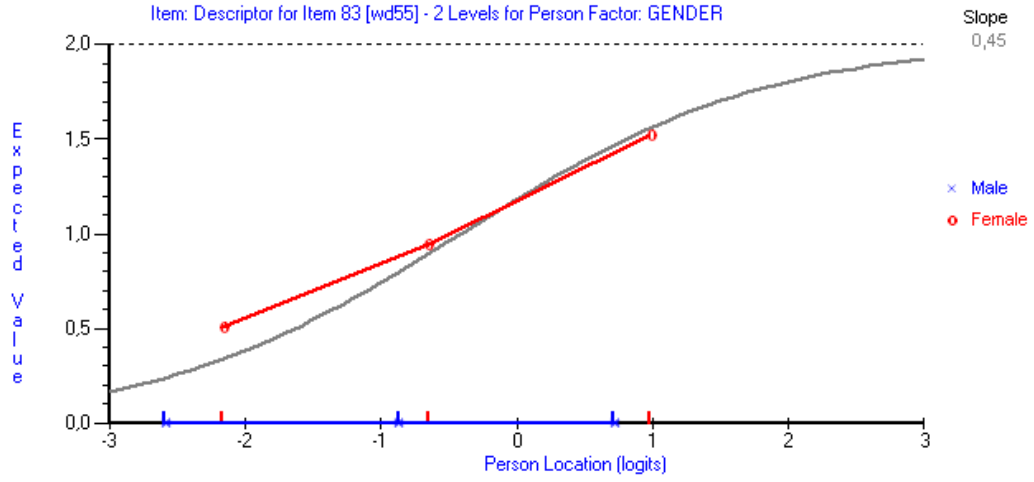
Şekil 3.8. “Aktivite-Ağrı” Soru Bankasındaki Maddelerin Eşik Değerleri



Şekil 3.8. (Devam) “Aktivite-Ağrı” Soru Bankasındaki Maddelerin Eşik Değerleri

Şekil 3.8 incelendiğinde soru bankasındaki maddelerin yanıt kategori sayıları eşit olmamakla beraber, eşik değerleri arasındaki uzaklık da birbirine eşit değildir. Bu durum da Rasch analizinde neden kısmi kredi modelinin kullanıldığını açık bir şekilde göstermektedir.

İkinci aşamada, soru bankasında yer alan 73 maddenin model uyumları incelenmiştir. Bunlar içerisinde v3, v13, v22, v24, wd25, wd31 ve h12 kodlu maddelerin ki-kare değerlerine ilişkin Bonferroni düzeltmeli olasılık (P) değerleri incelendiğinde modele uyumsuzluk gösterdikleri belirlenmiş, bu nedenle de soru bankasından çıkarılmışlardır. Bireylere ilişkin artık değerleri incelendiğinde modele uyumsuzluk gösteren birey bulunmamıştır. Bu aşamada 66 maddeden oluşan soru bankasının tek boyutlu olup olmadığı artıklar üzerinden yapılan TBA yardımıyla incelenmiştir. İlk temel bileşen üzerinde pozitif ve negatif yüklenen maddelerin ayrı birer boyut oluşturup oluşturmadığı test edildiğinde tek boyutluluk varsayımını bozan v1, v4, v6, v8, v16, v17, v18, wd23, wd52, wd53, wd54, h5, h6, h15, h18 ve h20 kodlu maddeler analizden çıkarılmıştır. Yerel bağımsızlık varsayımının sağlanıp sağlanmadığını test etmek için artık korelasyon matrisi incelenmiş, 0.50'nin üzerinde korelasyona sahip madde olmadığı için bu varsayımın sağlandığı belirlenmiştir. Cinsiyet ve yaş grubu bakımından MİF olup olmadığı grafiksel ve istatistiksel olarak değerlendirildiğinde yaş grubuna göre hiçbir madde MİF göstermezken; wd55 kodlu maddenin cinsiyete göre tekdüze MİF gösterdiği belirlenmiştir. Bu maddenin cinsiyete göre MKF eğrileri Şekil 3.9'da gösterilmiştir.



Şekil 3.9. “wd55” Kodlu Maddenin Cinsiyete Göre MKF Eğrileri

Şekil 3.9 incelendiğinde, wd55 kodlu “Son 30 gün içinde ev işlerini yeterince hızlı yapmada ne kadar güçlük çektiniz” maddesinin kadın hastalar için daha zor bir faaliyet olduğu sonucuna ulaşılır. Kadın hastaların, ev işi ile daha çok meşgul olmalarından dolayı erkek hastalara göre bu faaliyette daha fazla zorluk çekmeleri ve dolayısıyla bu madde için işlev farklılığının olması mantıklıdır. Bu madde, hastanın ev işleri ile ilgili özürülük düzeyini sorguladığından soru bankasında kalmasının gerekli olduğu düşünülmüştür. Ayrıca MİF varlığının denek sayısı ile bağlantılı olabileceği düşünülerek, denek sayısı artırılarak bu maddenin MİF gösterip göstermediğinin tekrar incelenmesine karar verilmiştir.

Bu sonuçlara göre elde edilen 50 maddelik “Aktivite-Ağrı” final soru bankasının tümel uyum iyiliği istatistikleri incelendiğinde madde etkileşim istatistiği ortalaması (standart sapması) -0.236 (0.861), birey etkileşim istatistiği ortalaması (standart sapması) -0.232 (0.743)'dir. Bu değerlerin ortalaması 0, standart sapması 1 değerine yakın olduğu için soru bankasında yer alan maddelerin ve bireylerin modele yeterli uyum gösterdiği söylenir. Madde- $\theta$  etkileşim istatistiği incelendiğinde ki-kare değeri (P değeri) 125.702 (0.042005) olduğundan ve Bonferroni düzeltmesine göre anlamlı olmadığından bu soru bankası için maddelere verilen yanıtların hiyerarşik

sıralamasının  $\theta$  düzeyi boyunca değişmediği ve değişmezlik özelliğinin sağlandığı sonucuna ulaşılır. Tümel uyum istatistiklerine ek olarak, “Aktivite-Ağrı” final soru bankasında yer alan her bir madde için artık değerleri ve ki-kare istatistiği cinsinden hesaplanan uyum istatistikleri Çizelge 3.7’de sunulmuştur.

Çizelge 3.7. “Aktivite-Ağrı” Final Soru Bankasındaki Maddelerin Model Uyum İstatistikleri

Maddeler	b	Standart Hata	Artık	Serbestlik Derecesi	Ki-Kare	Serbestlik Derecesi	Olasılık
v2	-3.173	0.153	-0.341	211.82	0.684	2	0.710
v5	-1.736	0.126	-0.581	212.79	3.738	2	0.154
v7	-0.649	0.120	1.296	212.79	3.540	2	0.170
v9	-2.419	0.137	0.269	212.79	0.931	2	0.628
v10	-1.888	0.129	0.400	212.79	0.746	2	0.689
v11	-1.354	0.125	-1.490	212.79	3.272	2	0.195
v12	-1.256	0.117	0.358	212.79	1.901	2	0.387
v14	-2.023	0.130	-1.213	212.79	1.290	2	0.525
v15	-1.210	0.124	-0.638	194.33	0.426	2	0.808
v19	-0.675	0.120	-0.748	211.82	0.891	2	0.640
v21	-0.698	0.120	0.352	210.85	0.945	2	0.623
n2	-1.657	0.169	-0.064	209.87	1.307	2	0.520
n8	-2.004	0.176	-0.825	209.87	0.050	2	0.975
n10	2.277	0.255	0.062	209.87	2.510	2	0.285
n11	-2.192	0.181	0.957	209.87	0.386	2	0.824
n14	3.837	0.444	-0.389	209.87	0.407	2	0.816
n17	-4.776	0.362	-0.590	209.87	0.552	2	0.759
n18	-2.554	0.192	-1.027	209.87	2.382	2	0.304
n19	-3.020	0.211	-0.309	209.87	1.002	2	0.606
n24	-3.378	0.232	-0.440	208.90	0.047	2	0.977
n25	-0.141	0.166	0.224	208.90	7.214	2	0.027
n26	-2.284	0.184	-0.844	208.90	1.722	2	0.423
n27	-3.551	0.243	-0.589	208.90	1.357	2	0.507
n28	-1.147	0.164	-0.544	208.90	2.453	2	0.293
n36	-5.316	0.450	-0.530	208.90	1.427	2	0.490
n38	-2.165	0.181	-0.278	208.90	2.197	2	0.333
wd21	-0.247	0.130	-0.341	212.79	1.338	2	0.512
wd22	0.917	0.151	0.227	212.79	1.505	2	0.471
wd24	1.369	0.148	0.153	212.79	0.898	2	0.638
wd32	4.159	0.177	-0.309	212.79	8.348	2	0.015
wd33	5.302	0.295	-0.291	212.79	0.369	2	0.831
wd34	1.614	0.153	1.491	207.93	1.971	2	0.373
wd55	-0.409	0.121	-0.165	195.30	0.078	2	0.962
wd61	0.694	0.123	1.664	212.79	6.951	2	0.031

Çizelge 3.7. (Devam) “Aktivite-Ağrı” Final Soru Bankasındaki Maddelerin Model Uyum İstatistikleri

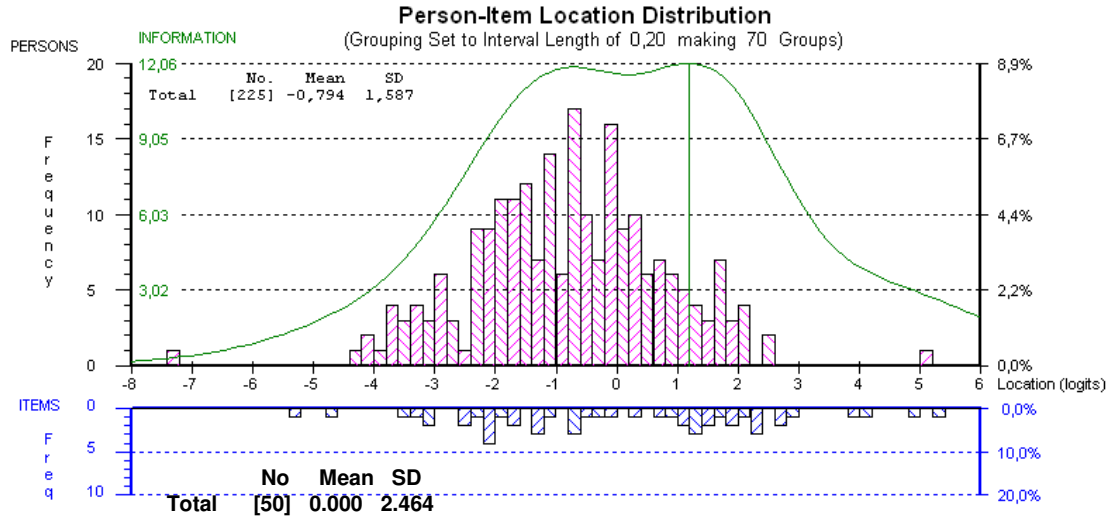
wd62	1.968	0.160	1.330	212.79	9.631	2	0.008
wd65	1.926	0.160	0.743	212.79	3.537	2	0.171
wd68	1.011	0.134	1.625	212.79	1.212	2	0.546
haq1	1.280	0.146	-0.221	208.90	2.128	2	0.345
haq2	2.244	0.174	0.065	208.90	4.617	2	0.099
haq3	2.829	0.162	-0.607	208.90	2.825	2	0.244
haq4	2.637	0.164	-1.623	208.90	5.601	2	0.061
haq7	2.323	0.184	-0.172	208.90	1.580	2	0.454
haq8	2.775	0.162	-2.660	208.90	10.304	2	0.006
haq9	1.202	0.167	-0.834	208.90	1.301	2	0.522
haq10	2.148	0.171	-1.448	207.93	3.571	2	0.168
haq13	0.274	0.122	0.073	206.96	0.345	2	0.842
haq14	1.197	0.150	-0.701	205.99	1.137	2	0.566
haq16	1.500	0.149	-0.351	207.93	5.556	2	0.062
haq17	4.923	0.248	-0.347	207.93	2.346	2	0.309
haq19	1.517	0.164	-1.595	207.93	5.173	2	0.075

Çizelge 3.7'ye göre artık değerleri  $\pm 2.5$  arasında değerler aldığından ve ki-kare değerlerine ilişkin olasılık değerleri Bonferroni düzeltmeli P değerinden ( $0.05/50=0.001$ ) daha yüksek olduğu için soru bankasındaki tüm maddelerin modele uyum gösterdiği söylenir.

Soru bankasının iç tutarlılığının incelenmesinde, veri setinde kayıp gözlem olduğu için, BAİ kullanılmış ve bu indeks 0.954 olarak elde edilmiştir. Bu değer yorumlandığında elde edilen soru bankasının güvenilir olduğu sonucuna ulaşılır.

“Aktivite-Ağrı” final soru bankasında yer alan maddelerin ve hastaların  $\theta$  düzeyi boyunca dağılımı Şekil 3.10'da verilmiştir.



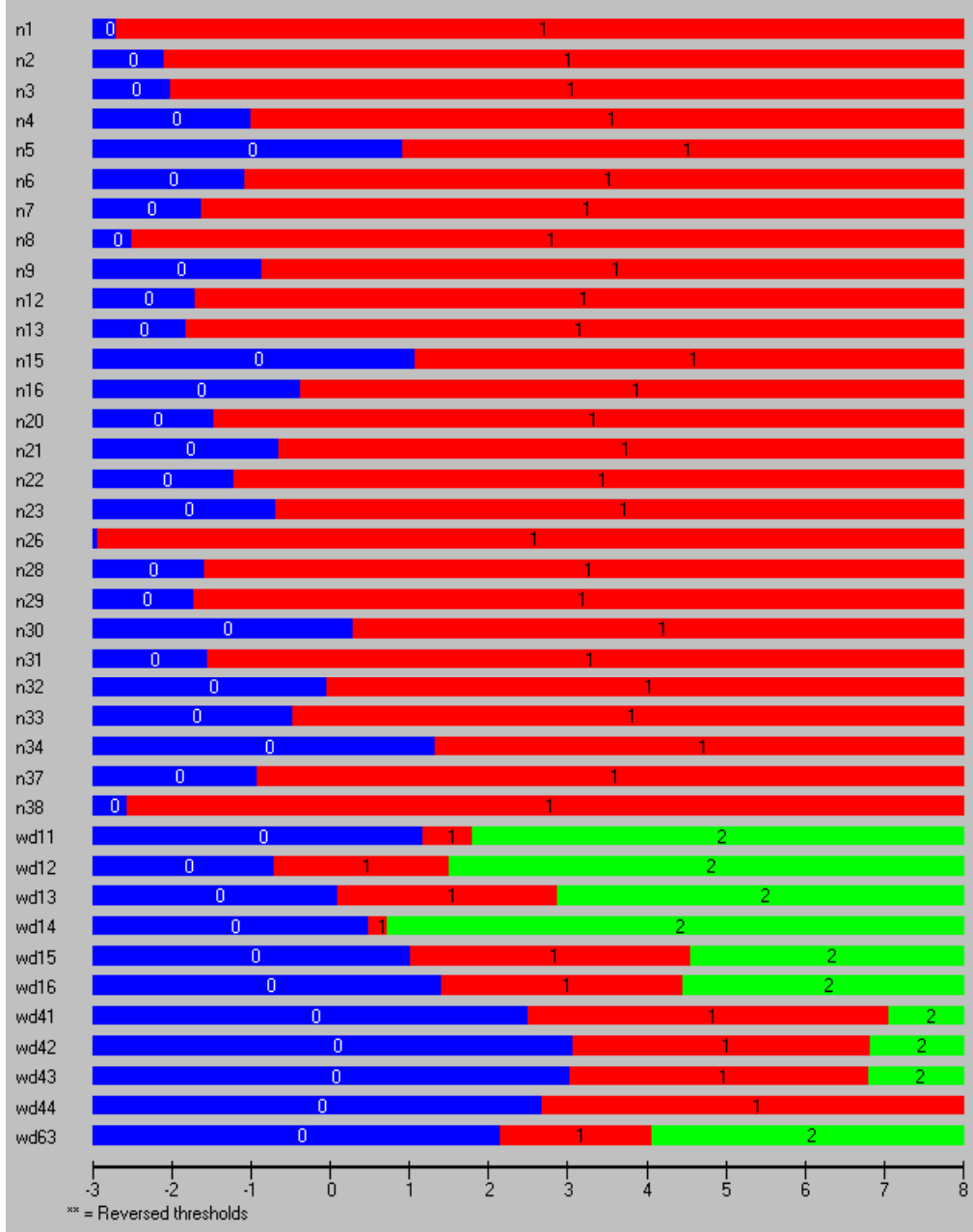


Şekil 3.10. “Aktivite-Ağrı” Final Soru Bankasındaki Maddelerin ve Hastaların  $\theta$  Düzeyi Boyunca Dağılımı

Soru bankasında yer alan maddelerin b değerlerinin ortalaması ile incelenen grubun ortalamasının birbirine uygun olması, dolayısıyla soru bankasının  $\theta$ 'nın tüm düzeylerinde bilgi verici olması istenir. Şekil 3.10'a göre, hem maddelerin, bireylerin konumları tarafından kapsandığı, hem de bu durumun tersinin doğru olduğu yani “Aktivite-Ağrı” final soru bankasının  $\theta$ 'nın tüm düzeylerinde yeterli sayıda madde içerdiği sonucuna ulaşılır. Soru bankasının  $\theta$  düzeyi -5.5 ile 5.5 arasında olan bireyler için daha fazla bilgi sağladığı söylenir.

### 3.2.4.2. “Ağrı-Sosyal-Bilişsel-Duygusal Durum” Soru Bankasının Rasch Analizi

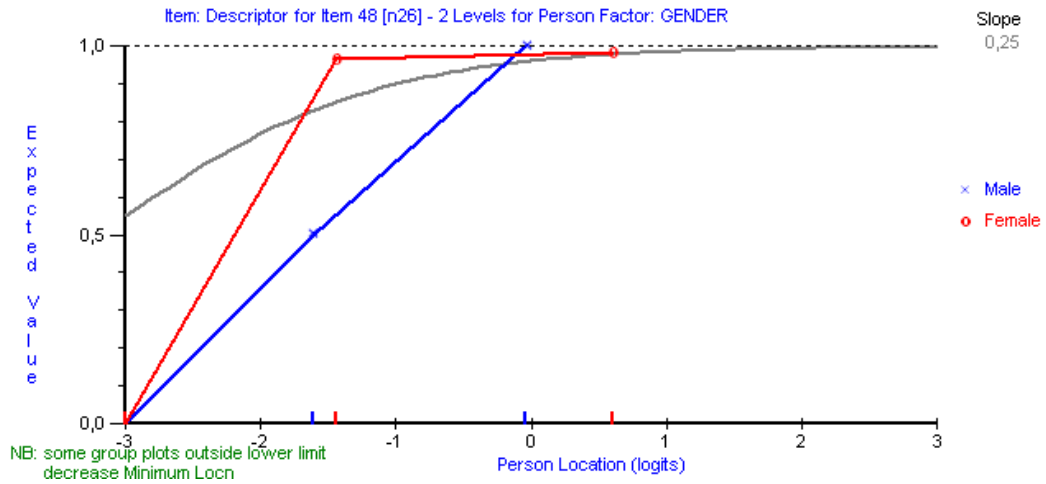
“Ağrı-Sosyal-Bilişsel-Duygusal Durum” soru bankasının Rasch analizinde ilk olarak, maddelere verilen yanıtlar, kategori olasılık eğrileri ile incelenmiş, eşik değerleri sıralı olmayan maddelerin bazı yanıt kategorileri birleştirilmiştir. Soru bankasındaki 38 madde için eşik değerlerine ilişkin birleştirme sonrasındaki yapı Şekil 3.11’de verilmiştir.



Şekil 3.11. “Ağrı-Sosyal-Bilişsel-Duygusal Durum” Soru Bankasındaki Maddelerin Eşik Değerleri

Şekil 3.11 incelendiğinde soru bankasındaki maddelerin yanıt kategori sayılarının yanında eşik değerleri arasındaki uzaklığın da birbirine eşit olmadığı görülmektedir.

İkinci aşamada, soru bankasında yer alan 38 maddenin model uyumları incelenmiştir. Bunlar içerisinde n3, n13 ve wd14 kodlu maddelerin ki-kare değerlerine ilişkin Bonferroni düzeltmeli olasılık (P) değerleri incelendiğinde modele uyumsuzluk gösterdikleri belirlenmiş, bu nedenle de soru bankasından çıkarılmışlardır. Bireylere ilişkin artık değerleri incelendiğinde modele uyumsuzluk gösteren birey bulunmamıştır. Bu aşamada 35 maddeden oluşan soru bankasının tek boyutlu olup olmadığı artıklar üzerinden yapılan TBA yardımıyla incelenmiştir. İlk temel bileşen üzerinde pozitif ve negatif yüklenen maddelerin ayrı birer boyut oluşturup oluşturmadığı test edildiğinde tek boyutluluk varsayımını bozan n32 ve n38 kodlu maddeler analizden çıkarılmıştır. Yerel bağımsızlık varsayımının sağlanıp sağlanmadığını test etmek için artık korelasyon matrisi incelenmiş, 0.50'nin üzerinde korelasyona sahip madde olmadığı için bu varsayımın sağlandığı belirlenmiştir. Cinsiyet ve yaş grubu bakımından MİF olup olmadığı grafiksel ve istatistiksel olarak değerlendirildiğinde yaş grubuna göre hiçbir madde MİF göstermezken; n26 kodlu maddenin cinsiyete göre tekdüze MİF gösterdiği tespit edilmiştir. Bu maddenin cinsiyete göre MKF eğrileri Şekil 3.12'de gösterilmiştir.



Şekil 3.12. “n26” Kodlu Maddenin Cinsiyete Göre MKF Eğrileri

Şekil 3.12 incelendiğinde, “n26” kodlu “Çabucak yoruluyorum” maddesinin kadın hastalar tarafından daha zor bir faaliyet olduğu görülmektedir. Bu madde, hastanın fiziksel açıdan özürülük düzeyini sorguladığından soru bankasında kalmasının gerekli olduğu düşünülmüştür. Ayrıca MİF varlığının denek sayısı ile bağlantılı olabileceği düşünülerek, denek sayısının artırılarak bu maddenin MİF gösterip göstermediğinin tekrar incelenmesine karar verilmiştir.

Bu sonuçlara göre elde edilen 33 maddelik “Ağrı-Sosyal-Bilişsel-Duygusal Durum” final soru bankasının tümel uyum iyiliği istatistikleri incelendiğinde madde etkileşim istatistiği ortalaması (standart sapması) - 0.331 (1.059), birey etkileşim istatistiği ortalaması (standart sapması) -0.391 (0.890)'dir. Bu değerlerin ortalaması 0, standart sapması 1 değerine yakın olduğu için soru bankasında yer alan maddelerin ve bireylerin modele yeterli uyum gösterdiği söylenir. Madde- $\theta$  etkileşim istatistiği incelendiğinde ki-kare değeri (P değeri) 94.287 (0.012748) olduğundan ve Bonferroni düzeltmesine göre anlamlı olmadığından bu soru bankası için değişmezlik özelliğinin sağlandığı yorumuna ulaşılır. Tümel uyum istatistiklerine ek olarak, “Aktivite-Ağrı” final soru bankasında yer alan her bir madde için artık değerleri ve ki-kare istatistiği cinsinden hesaplanan uyum istatistikleri Çizelge 3.8'de sunulmuştur.

Çizelge 3.8. “Ağrı-Sosyal-Bilişsel-Duygusal Durum” Final Soru Bankasındaki Maddelerin Model Uyum İstatistikleri

Maddeler	b	Standart Hata	Artık	Serbestlik Derecesi	Ki-Kare	Serbestlik Derecesi	Olasılık
n1	-2.921	0.188	0.504	202.290	6.544	2	0.038
n2	-2.306	0.176	1.531	202.290	3.654	2	0.161
n4	-1.196	0.171	0.418	202.290	0.302	2	0.860
n5	0.736	0.202	0.371	202.290	3.979	2	0.137
n6	-1.283	0.171	0.753	202.290	0.227	2	0.893
n7	-1.830	0.172	-0.823	202.290	7.291	2	0.026
n8	-2.738	0.184	-0.293	202.290	1.268	2	0.531
n9	-1.064	0.172	-1.050	202.290	2.227	2	0.328
n12	-1.921	0.172	-1.124	202.290	3.149	2	0.207
n15	0.888	0.207	-1.428	202.290	1.537	2	0.464
n16	-0.568	0.175	-1.716	202.290	2.478	2	0.290

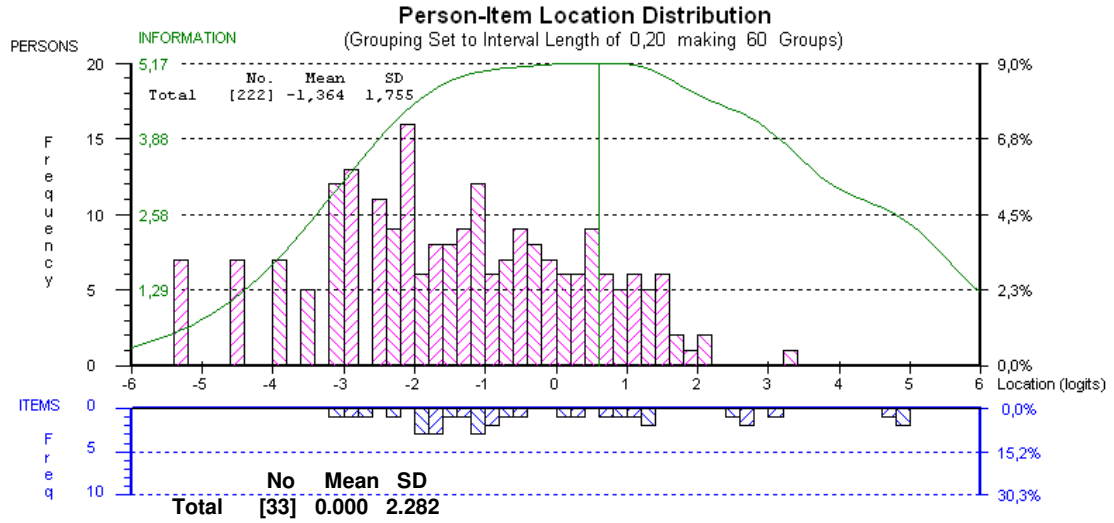
Çizelge 3.8. (Devam) “Ağrı-Sosyal-Bilişsel-Duygusal Durum” Final Soru Bankasındaki Maddelerin Model Uyum İstatistikleri

Maddeler	b	Standart Hata	Artık	Serbestlik Derecesi	Ki-Kare	Serbestlik Derecesi	Olasılık
n20	-1.682	0.171	0.135	202.290	2.288	2	0.319
n21	-0.841	0.173	-1.135	202.290	3.005	2	0.223
n22	-1.430	0.171	-0.699	202.290	1.049	2	0.592
n23	-0.866	0.173	-1.906	201.330	3.886	2	0.143
n26	-3.194	0.197	-1.090	201.330	4.604	2	0.100
n28	-1.787	0.172	0.115	201.330	1.373	2	0.503
n29	-1.949	0.173	-0.370	201.330	1.835	2	0.400
n30	0.103	0.186	-0.974	201.330	4.205	2	0.122
n31	-1.767	0.172	-1.390	201.330	1.904	2	0.386
n33	-0.668	0.175	-1.515	201.330	5.154	2	0.076
n34	1.156	0.218	-0.920	201.330	1.394	2	0.498
n37	-1.114	0.172	-0.680	201.330	2.690	2	0.260
wd11	1.347	0.157	2.203	206.150	8.113	2	0.017
wd12	0.232	0.135	1.536	206.150	7.348	2	0.025
wd13	1.327	0.158	0.168	205.180	1.213	2	0.545
wd15	2.626	0.194	0.002	206.150	1.350	2	0.509
wd16	2.741	0.206	1.645	206.150	0.521	2	0.771
wd41	4.687	0.298	-1.541	202.290	2.419	2	0.298
wd42	4.876	0.363	-1.004	205.180	1.144	2	0.564
wd43	4.821	0.351	-0.884	206.150	1.246	2	0.536
wd44	2.549	0.319	0.637	200.370	3.889	2	0.143
wd63	3.035	0.244	-0.399	206.150	1.000	2	0.606

Çizelge 3.8’e göre artık değerleri  $\pm 2.5$  arasında değerler aldığından ve ki-kare değerlerine ilişkin olasılık değerleri Bonferroni düzeltmeli P değerinden ( $0.05/33=0.0015$ ) daha yüksek olduğu için soru bankasındaki tüm maddelerin modele uyum gösterdiği söylenir.

Soru bankasının iç tutarlılığının incelenmesinde, veri setinde kayıp gözlem olduğu için, BAİ kullanılmış ve bu indeks 0.920 olarak elde edilmiştir. Bu değer yorumlandığında elde edilen soru bankasının güvenilir olduğu sonucuna ulaşılır.

“Ağrı-Sosyal-Bilişsel-Duygusal Durum” final soru bankasında yer alan maddelerin ve hastaların  $\theta$  düzeyi boyunca dağılımı Şekil 3.13’te verilmiştir.



Şekil 3.13. “Ağrı-Sosyal-Bilişsel-Duygusal Durum” Final Soru Bankasındaki Maddelerin ve Hastaların  $\theta$  Düzeyi Boyunca Dağılımı

Soru bankasında yer alan maddelerin b değerlerinin ortalaması ile incelenen grubun ortalamasının birbirine uygun olması, dolayısıyla soru bankasının  $\theta$ 'nın tüm düzeylerinde bilgi verici olması istenen bir durumdur. Şekil 3.13'e göre, maddelerin bireylerin konumları tarafından kapsandığı, fakat bu durumun tersinin doğru olmadığı yani “Ağrı-Sosyal-Bilişsel-Duygusal Durum” final soru bankasının  $\theta$ 'nın tüm düzeylerinde yeterli sayıda madde içermediği sonucuna ulaşılır. Bu yüzden, soru bankasının  $\theta$  düzeyi -3 ile 5 arasında olan bireyler için daha fazla bilgi sağladığı söylenir.

### 3.2.5. Benzetim Çalışması ve Gerçek Uygulama Sonuçları

Benzetim çalışmasında, Rasch analizi sonucunda elde edilen madde parametreleri kullanılarak RUMMss programı ile 10000 hastanın yanıtları türetilmiştir. Aynı zamanda bu yanıt yapıları ile hastaların özürlülük düzeyleri ( $\theta_{Rasch}$ ) de elde edilmiştir. RUMMss programından türetilen yanıtlarla CATSIM modülü kullanılarak, 10000 hastanın özürlülük düzeyi ( $\theta_{BUT}$ ) BUT yöntemiyle de bulunmuştur.  $\theta_{Rasch}$  ile  $\theta_{BUT}$  değerleri Blant-Altman yöntemi, Spearman ilişki katsayısı ve SKK ile karşılaştırılmıştır. Gerçek uygulamada

ise, elde edilen final soru bankalarında yer alan maddeler, 43 hastaya klasik yöntem ve BUT yöntemi ile uygulanmış; her iki uygulama sonucunda hastalara ilişkin elde edilen özürlülük düzeyleri (sırasıyla  $\theta_{Raschg}$  ve  $\theta_{BUTg}$ ) benzetim çalışmasında kullanılan yöntemlerle karşılaştırılmıştır. Bu bölümde, elde edilen sonuçlar soru bankaları için ayrı ayrı ele alınacaktır.

### **3.2.5.1. “Aktivite-Ağrı” Soru Bankası için Benzetim Çalışması ve Gerçek Uygulama Sonuçları**

#### *Benzetim Çalışması*

“Aktivite-Ağrı” final soru bankası ile yapılan BUT uygulamasında 10000 hastanın 9408'i (%94.1) için özürlülük düzeyi belirlenebilirken; 11'i (%0.1) için en düşük özürlülük düzeyi ataması yapılmış (11 hasta BUT sürecinde tüm maddelere en düşük kategoride yanıt vermişlerdir), 581 (%5.8) hasta için de Newton-Raphson algoritmasından kaynaklanan sorunlar nedeniyle özürlülük düzeyi elde edilememiştir. Bu nedenle “Aktivite-Ağrı” soru bankasına ilişkin karşılaştırmalar, 9419 (%94.2) hasta üzerinden yapılmıştır.

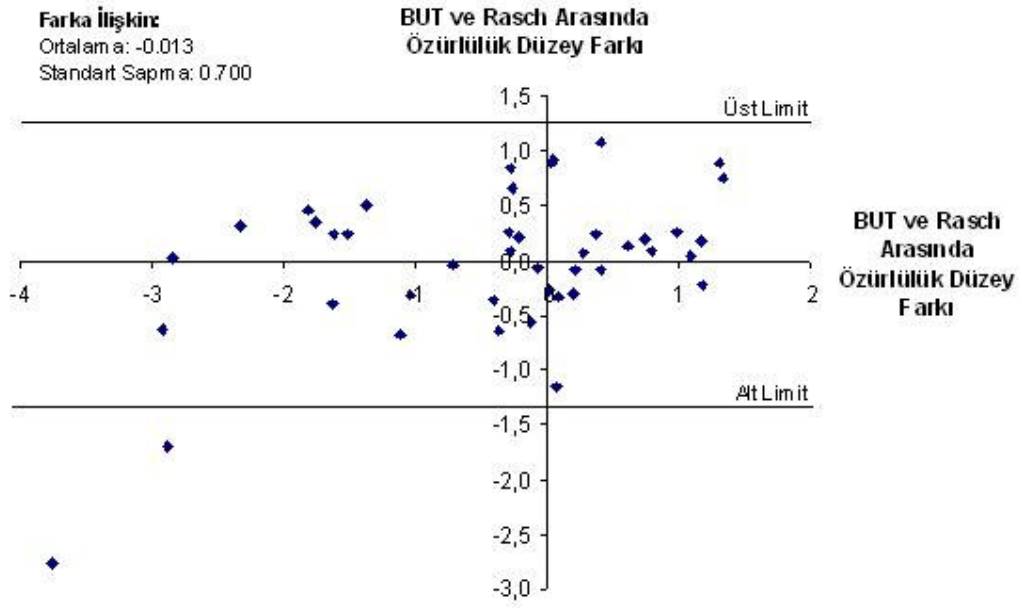
Elde edilen sonuçlar Blant-Altman yöntemi ile karşılaştırıldığında, “Aktivite-Ağrı” final soru bankası için  $\theta_{Rasch}$  ve  $\theta_{BUT}$  değerleri arasındaki farkların %95'i (-1.420, 1.280) uyum sınırları içerisindedir. Uyum sınırları incelendiğinde BUT yöntemi ile elde edilen özürlülük düzeylerinin %5'inin, Rasch analizi ile elde edilenlerden 1.420 lojit daha düşük ya da 1.280 lojit daha yüksek bulunabileceği yorumuna ulaşılır. Buna göre 9419 hastanın 8975'i (%95.3) uyum sınırları içerisinde kalırken; 444 (%4.7) hastanın 257'sinde BUT yöntemi ile elde edilen; 187'sinde Rasch analizi ile elde edilen özürlülük düzeyi daha yüksek bulunmuştur.

$\theta_{\text{Rasch}}$  ile  $\theta_{\text{BUT}}$  deęerleri Spearman iliřki katsayısı ile karřılařtırıldıęında, iki deęer arasında pozitif ynl bir iliřki belirlenmiřtir ( $r=0.97$ ,  $p<0.001$ ). Aynı deęerler iin SKK 0.94 ( $p<0.001$ ) olarak bulunmuřtur. Bu katsayı iin 0.85-0.94 arasındaki deęerlerin, yksek uyumu gsterdięi bilgisi iřıęında Rasch analizi ile elde edilen zrllk dzeyleri ile BUT yntemi ile elde edilenlerin ok uyumlu olduęu sonucuna ulařılır. Bireylerin zrllk dzeyleri, RUMMss programında soru bankasındaki maddelerin tm kullanılarak; CATSIM programında ise BUT yntemi kullanılarak elde edildięinden yukarıdaki sonular madde sayısı bakımından da deęerlendirilmiřtir. Buna gre  $\theta_{\text{Rasch}}$  deęerleri 50 madde zerinden elde edilirken;  $\theta_{\text{BUT}}$  deęerleri ortalama 15 (standart sapma: 3, minimum-maksimum: 5-50) madde kullanılarak bulunmuřtur.

### Gerek Uygulama

“Aktivite-Aęrı” final soru bankası ile gerek hastalar zerinde yapılan BUT uygulamasında  $\theta_{\text{Rasch}_g}$  ve  $\theta_{\text{BUT}_g}$  deęerleri arasındaki farkların %95’i (-1.385, 1.360) Blant-Altman uyum sınırları ierisindedir. Uyum sınırları incelendięinde BUT yntemi ile elde edilen zrllk dzeylerinin %5’inin, Rasch analizi ile elde edilenlerden 1.385 lojit daha dřk ya da 1.360 lojit daha yksek bulunabileceęi yorumuna ulařılır. Buna gre 43 hastanın 41’i (%95) uyum sınırları ierisinde kalırken; 2 hasta (%5) iin Rasch analizi ile elde edilen zrllk dzeyi daha yksek bulunmuřtur. Gerek uygulama sonularına gre “Aktivite-Aęrı” final soru bankası iin uyum sınırlarına iliřkin Blant-Altman grafięi Őekil 3.14’te verilmiřtir.





Şekil 3.14. Gerçek Uygulama Sonuçlarına Göre “Aktivite-Ağrı” Final Soru Bankası İçin Uyum Sınırlarına İlişkin Blant-Altman Grafiği

$\theta_{Raschg}$  ile  $\theta_{BUTg}$  değerleri Spearman ilişki katsayısı ile karşılaştırıldığında, iki değer arasında pozitif yönlü bir ilişki belirlenmiştir ( $r=0.87$ ,  $p<0.001$ ). Aynı değerler için SKK 0.86 ( $p<0.001$ ) olarak bulunmuştur. Bu katsayı için 0.85-0.94 arasındaki değerlerin, yüksek uyumu gösterdiği bilgisi ışığında Rasch analizi ile elde edilen özürülük düzeyleri ile BUT yöntemi ile elde edilenlerin çok uyumlu olduğu sonucuna ulaşılır.  $\theta_{Raschg}$  değerleri soru bankasındaki maddelerin tümü kullanılarak; CAT programında ise BUT yöntemi kullanılarak elde edildiğinden yukarıdaki sonuçlar madde sayısı bakımından da değerlendirilmiştir. Buna göre  $\theta_{Raschg}$  değerleri 50 madde üzerinden elde edilirken;  $\theta_{BUTg}$  değerleri ortalama 16 (standart sapma: 3, minimum-maksimum: 5-19) madde kullanılarak bulunmuştur.

### 3.2.5.2. “Ağrı-Sosyal-Bilişsel-Duygusal Durum” Soru Bankası için Benzetim Çalışması ve Gerçek Uygulama Sonuçları

#### Benzetim Çalışması

“Ağrı-Sosyal-Bilişsel-Duygusal Durum” final soru bankası ile yapılan BUT uygulamasında 10000 hastanın 8216’sı (%82.1) için özürlülük düzeyi belirlenebilirken; 147’si (%1.5) için en düşük özürlülük düzeyi ataması yapılmış (147 hasta BUT sürecinde tüm maddelere en düşük kategoride yanıt vermişlerdir), 1637 (%16.4) hasta için de Newton-Raphson algoritmasından kaynaklanan sorunlar nedeniyle özürlülük düzeyi elde edilememiştir. Bu nedenle “Ağrı-Sosyal-Bilişsel-Duygusal Durum” soru bankasına ilişkin karşılaştırmalar, 8363 (%83.6) hasta üzerinden yapılmıştır.

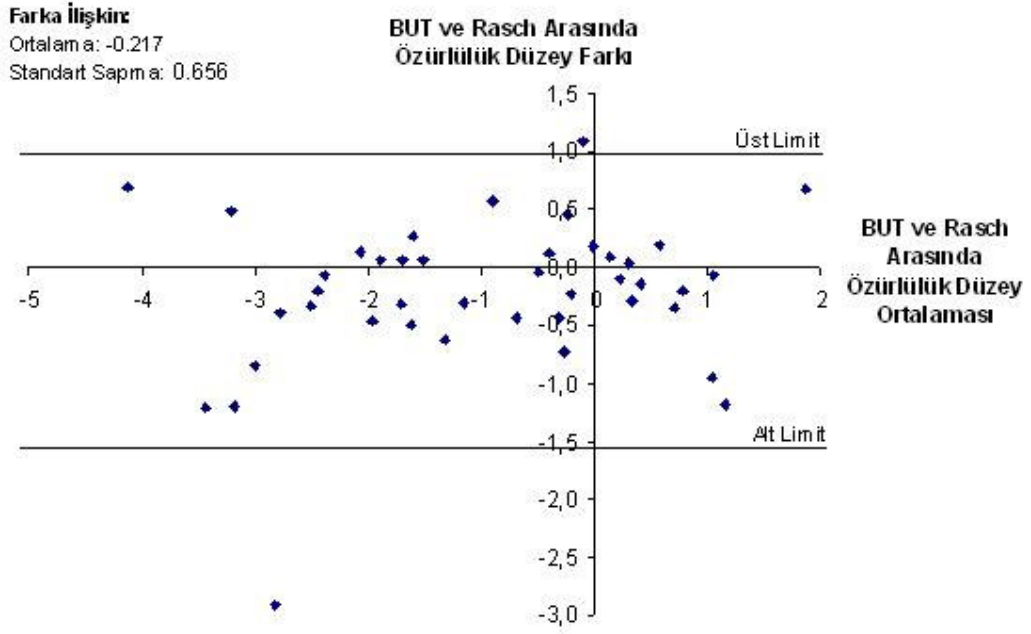
Elde edilen sonuçlar Blant-Altman yöntemi ile karşılaştırıldığında, “Ağrı-Sosyal-Bilişsel-Duygusal Durum” final soru bankası için  $\theta_{\text{Rasch}}$  ve  $\theta_{\text{BUT}}$  değerleri arasındaki farkların %95’i (-1.747, 1.236) uyum sınırları içerisindedir. Uyum sınırları incelendiğinde BUT yöntemi ile elde edilen özürlülük düzeylerinin %5’inin, Rasch analizi ile elde edilenlerden 1.747 lojit daha düşük ya da 1.236 lojit daha yüksek bulunabileceği yorumuna ulaşılmıştır. Buna göre 8363 hastanın 7927’si (%94.8) uyum sınırları içerisinde kalırken; 436 (%5.2) hastanın 168’inde BUT yöntemi ile elde edilen; 268’inde Rasch analizi ile elde edilen özürlülük düzeyi daha yüksek bulunmuştur.

$\theta_{\text{Rasch}}$  ile  $\theta_{\text{BUT}}$  değerleri Spearman ilişki katsayısı ile karşılaştırıldığında, iki değer arasında pozitif yönlü bir ilişki belirlenmiştir ( $r=0.95$ ,  $p<0.001$ ). Aynı değerler için SKK 0.91 ( $p<0.001$ ) olarak bulunmuştur. Bu katsayı için 0.85-0.94 arasındaki değerlerin, yüksek uyumu gösterdiği bilgisi ışığında Rasch analizi ile elde edilen özürlülük düzeyleri ile BUT yöntemi ile elde edilenlerin çok uyumlu olduğu sonucuna ulaşılabilir. Bireylerin özürlülük düzeyleri,

RUMMss programında soru bankasındaki maddelerin tümü kullanılarak; CATSIM programında ise BUT yöntemi kullanılarak elde edildiğinden yukarıdaki sonuçlar madde sayısı bakımından da değerlendirilmiştir. Buna göre  $\theta_{\text{Rasch}}$  değerleri 33 madde üzerinden elde edilirken;  $\theta_{\text{BUT}}$  değerleri ortalama 16 (standart sapma: 4, minimum-maksimum: 6-33) madde kullanılarak bulunmuştur.

### Gerçek Uygulama

“Ağrı-Sosyal-Bilişsel-Duygusal Durum” final soru bankası ile gerçek hastalar üzerinde yapılan BUT uygulamasında  $\theta_{\text{Rasch}_g}$  ve  $\theta_{\text{BUT}_g}$  değerleri arasındaki farkların %95’i (-1.503, 1.069) Blant-Altman uyum sınırları içerisindedir. Uyum sınırları incelendiğinde BUT yöntemi ile elde edilen özürlülük düzeylerinin %5’inin, Rasch analizi ile elde edilenlerden 1.503 lojit daha düşük ya da 1.069 lojit daha yüksek bulunabileceği yorumuna ulaşılır. Buna göre 43 hastanın 41’i (%95) uyum sınırları içerisinde kalırken; 2 hastanın (%5) 1’inde BUT yöntemi ile diğerinde Rasch analizi ile elde edilen özürlülük düzeyi daha yüksek bulunmuştur. Gerçek uygulama sonuçlarına göre “Ağrı-Sosyal-Bilişsel-Duygusal Durum” final soru bankası için uyum sınırlarına ilişkin Blant-Altman grafiği Şekil 3.15’te verilmiştir.



Şekil 3.15. Gerçek Uygulama Sonuçlarına Göre “Ağrı-Sosyal-Bilişsel-Duygusal Durum” Final Soru Bankası İçin Uyum Sınırlarına İlişkin Blant-Altman Grafiği

$\theta_{Raschg}$  ile  $\theta_{BUTg}$  değerleri Spearman ilişki katsayısı ile karşılaştırıldığında, iki değer arasında pozitif yönlü bir ilişki belirlenmiştir ( $r=0.94$ ,  $p<0.001$ ). Aynı değerler için SKK 0.91 ( $p<0.001$ ) olarak bulunmuştur. Bu katsayı için 0.85-0.94 arasındaki değerlerin, yüksek uyumu gösterdiği bilgisi ışığında Rasch analizi ile elde edilen özürülük düzeyleri ile BUT yöntemi ile elde edilenlerin çok uyumlu olduğu sonucuna ulaşılır.  $\theta_{Raschg}$  değerleri soru bankasındaki maddelerin tümü kullanılarak; CAT programında ise BUT yöntemi kullanılarak elde edildiğinden yukarıdaki sonuçlar madde sayısı bakımından da değerlendirilmiştir. Buna göre  $\theta_{Raschg}$  değerleri 33 madde üzerinden elde edilirken;  $\theta_{BUTg}$  değerleri ortalama 17 (standart sapma: 3, minimum-maksimum: 6-23) madde kullanılarak bulunmuştur.

## 4. TARTIŞMA

Elde edilen bulgular, bel ağrılı ve diz osteoartritli hastalar için ICF sınıflamasını temel alarak geliştirilen soru bankalarının psikometrik özelliklerinin yeterli olduğunu; soru bankalarında yer alan maddelerin, bu hastaların özürölülük düzeylerinin belirlenmesi amacıyla BUT uygulamalarında kullanılabileceğini; BUT uygulaması sonucunda hastaların özürölülük düzeylerinin klasik uygulama sonucunda elde edilen bulgularla uyumlu olarak daha az madde ile daha kısa sürede belirlenebildiğini göstermiştir.

Çalışmanın başlangıç aşamasında yapılan literatür incelemesi sonucunda, soru bankalarının oluşturulmasında tek bir ölçekten elde edilen maddeler yerine farklı ölçeklerden elde edilen maddelerin kullanıldığı belirlenmiştir. Bizim çalışmamızda olduğu gibi, Davidson ve ark. (2004) tarafından yapılan bir çalışmada da, bel ağrılı hastaların fonksiyonellik düzeylerinin değerlendirilmesinde SF-36 (Short Form Health Survey) fiziksel işlerlik ölçeğinden 10 madde, Oswestry ve Quebec bel ağrısı ölçeklerinden de 4'er madde olmak üzere toplam 18 maddelik yeni bir ölçek geliştirilmiştir. Yine başka bir çalışmada romatoid artritli hastalarda tedaviye cevabın değerlendirilmesinde modifiye HAQ ve SF-36'nın fiziksel bölümü birleştirilerek yeni bir ölçek elde edilmiştir (Martin ve ark., 2007). Bu çalışmalarda da belirtildiği gibi, mevcut ölçeklerin birleştirilmesi ile tedaviye bağlı değişimlerin daha güçlü olarak ortaya konduğunu ve soru bankasının ölçeğin orijinallerine göre daha hassas olduğunu düşünmekteyiz.

Soru bankalarının ICF çekirdek dizinine göre belirlenen maddelerden oluşması, elde edilecek sonuçların standartlaştırılması açısından çok önemlidir. Bu nedenle çalışmamızda maddelerin ICF sınıflamasına göre eşleştirilmesinde Cieza ve ark. (2002) tarafından yapılan çalışma kılavuz olarak kullanılmıştır. Dolayısıyla çalışmamız, soru bankalarındaki maddelerin

ICF tarafından belirlenen sınıflamalarla örtüşecek şekilde oluşturulması bakımından benzerlerine üstünlük sağlamaktadır.

Bjorner ve ark. (2003) tarafından yapılan çalışmada, baş ağrısının etkilerinin değerlendirilmesi için geliştirilen soru bankasının faktör analizi sonuçlarının, YHKOK uyum istatistiği incelendiğinde tatmin edici olmadığı belirtilmiştir. Aynı çalışmada, sürekli değişkenlerden oluşan veri setleri için YHKOK değerinin 0.05'in altında olması durumunda iyi uyum olduğu biçiminde standart bir kural olduğu halde, kategorik değişkenler için uyum istatistiklerinin dağılımı, yani uygun kesim noktası hakkında yeterli bilgi bulunmadığı ifade edilmiştir. Bu nedenle, model uyumunu arttırmak için bazı düzenlemeler yapıldığı fakat çok fazla bir gelişme sağlanamadığı, ancak faktör yüklerinin ve ilişki katsayılarının büyüklüğünün yorumlanabilir olduğu gösterilmiştir. Faktör analizi, Rasch analizi sonucunda belirlenecek final modele ulaşmak için bir aşama olarak değerlendirildiğinden, elde edilen sonuçların yine de yeterli olduğu yorumu yapılmıştır. Fliege ve ark. (2005) tarafından bireylerde depresyon düzeyinin belirlenmesi için BUT yönteminin uygulandığı diğer bir çalışmada, uyum istatistikleri çok iyi olmasa da Tip I Hata ve Tip II hata risklerini dengelemek için, analizler sonucunda elde edilen modellerin yeterli olduğu kabul edilmiştir. Bizim çalışmamızda da soru bankalarında yer alan maddelerin faktör yapısını ortaya çıkarmak için kategorik verilerde AFA ve DFA analizleri yapılmış, ancak elde edilen uyum istatistiklerinin kabul edilebilir olup olmadığına ilişkin net bir çıkarsama yapılamamıştır. Bu nedenle sonuçların değerlendirilmesinde, Bjorner ve ark. (2003) tarafından önerilen yaklaşım benimsenerek model uyumunun yeterli olduğu sonucuna ulaşılmıştır.

Çalışmamızda soru bankalarında yer alan maddelerin iki boyutta toplandığı belirlendiğinden soru bankalarının psikometrik özellikleri her bir boyut bazında incelenmiştir. Kopec ve ark. (2006) tarafından artritli hastalarda sağlığa ilişkin yaşam kalitesinin belirlenmesi amacıyla yapılan bir çalışmada bu alanda yaygın olarak kullanılan ölçeklerden sağlıkla ilişkili

yaşam kalitesinin 5 farklı alt boyutu saptanmış; her bir boyut için soru bankası geliştirilmesinin gerekliliği ve bu bankaların geliştirilmesi için de MYT modellerinden yararlanılması gerektiği üzerinde önemle durulmuştur.

Tek boyutlu bir değişken oluşturmak için gerekli ve yeterli ölçüm modellerinin Rasch türü modeller olduğu gösterildiğinden, BUT uygulamalarında genellikle bu psikometrik modeller üzerine odaklanılmıştır. Diğer psikometrik modeller önceden var olan soru bankalarının özelliğinden dolayı seçilse de, BUT uygulamaları için soru bankasının geliştirilmesi üzerindeki kısıtlılıklar, Rasch modellerinin kullanılmasını gerektirir (Linacre, 2000). Sağlık alanında yapılan çalışmalarda yaygın olarak kullanılan Rasch türü modellerde (McHorney ve Monahan, 2004), maddeler b değerleri bakımından farklılık gösterirken; eşit ayırt ediciliğe sahip oldukları varsayılır (Rasch, 1980; Wright ve Masters, 1982; Hays ve ark., 2000). Çalışmamızda soru bankasında hem iki sonuçlu hem de çok sonuçlu maddeler yer aldığından ve eşik değerlerine ilişkin grafikler incelendiğinde eşik değerleri arasındaki uzaklıklar eşit olmadığından Rasch türü modellerden Kısmi Kredi Modeli kullanılmıştır. Hart ve ark. (2005) tarafından yapılan çalışmada da soru bankasında yer alan maddeler çok sonuçlu olduğundan ve eşik değerleri arasındaki uzaklıklar eşit olduğundan, maddelerin kalibrasyonu için Sıralı Sonuçlu Modelin kullanıldığı belirtilmiştir.

Çalışmamızda, çok sonuçlu maddelerin bazılarının yanıt kategorilerine ilişkin eşik değerlerinin sıra takip etmediği belirlendiğinden, bazı yanıt kategorileri birleştirilerek bu sorun çözümlenmiştir. Bu durumun hastaların yaşlı ve düşük eğitim düzeyine sahip olmaları nedeniyle, yanıt kategorileri arasında ayrımsama yapmada zorluk çekmelerinden kaynaklandığı düşünülmektedir.

Soru bankasının geçerliđi, sahip olduđu psikometrik özelliklere bađlıdır. Ancak, soru bankasının istenilen psikometrik özelliklerin tümüne sahip olması her zaman mümkün olmamaktadır. Bunun nedenleri arasında, soru bankasında yer alan maddelerin farklı ölçeklerden gelmesi nedeniyle,

- Maddelerin yanıt kategori sayılarının farklı olması,
- Maddeler arasında yüksek derecede ilişki olması,
- Maddelerin tümünün tek bir özelliđi ölçmemesi,
- Bazı maddelerin işlevlerinin gruplar için farklılık göstermesi,
- Maddelerin MYT altında yer alan modellere uyum göstermemesi

olduđu düşünölmektedir. Çalışmamızda, soru bankalarının psikometrik özellikleri “Gereç ve Yöntem” bölümünde bahsedilen protokol doğrultusunda incelenmiş, modele uyumsuzluk gösteren, tek boyutluluk veya yerel bağımsızlık varsayımını bozan fazla sayıda madde olmadığı belirlenmiştir. Model uyumunu incelemede, çoklu karşılaştırma için Bonferroni düzeltmesi uygulanmıştır (Blant ve Altman, 1995). Soru bankalarında yer alan maddelerin önemli ölçüde MIF’e sahip olmadıkları görölmüştür. Sonuç olarak elde edilen soru bankalarının BAI değerleri de oldukça yüksek olduğundan, bel ađrılı ve diz osteoartritli hastaların özörlölük değerlendirmesinde güvenilir bir biçimde kullanılabilecekleri tespit edilmiştir.

Soru bankasında  $\theta$ ’nın tüm düzeylerinde yeterli sayıda madde olduğunda, BUT uygulamasından elde edilen sonuçlar daha güvenilir olacaktır. Ancak, sosyal bilimler ve eğitim bilimlerinde olduđu gibi sađlık alanında yapılan çalışmalarda madde sayısını arttırmak her zaman mümkün olmamaktadır. Çalışmamızda, her bir hastalık grubu için oluşturulan soru bankalarında  $\theta$ ’nın bazı düzeylerinde yeterli sayıda madde bulunmadığından, ileride yapılacak benzer amaçlı çalışmalar için  $\theta$ ’nın bu düzeylerine uygun yeni maddelerin soru bankalarına eklenmesi önerilmektedir. Bjorner ve ark. (2003) tarafından yapılan çalışmada da, mevcut ölçeklerden elde edilen soru



bankasının, en fazla bilgiyi  $\theta$  düzeyi yüksek olan hastalar için sağladığı, bu nedenle de orjinal ölçeklerin sahip olduğu bazı yetersizliklere sahip olduğu belirtilmiştir.

Yaygın olarak kullanılan sonuç değerlendirim ölçeklerinin çoğu, hastaların sonuçlarının izlenmesinde ve değerlendirilmesinde gerekli olan güvenilirlik ve kesinlik ölçümleri bakımından yetersiz kalabilmektedir. Bu soruna bir çözüm, güvenilirliği ve ölçüm kesinliğini arttırmak için daha fazla maddeden oluşan ölçme araçlarının kullanılmasıdır. Bununla beraber bu yaklaşım, zamanın sınırlı olduğu klinik uygulamalarda hasta bıkkınlığına da yol açarak bir çözüm niteliği taşımayacaktır. Ölçüm kesinliği ve hasta bıkkınlığı arasında kurduğu dengeden dolayı en uygun çözüm, BUT yönteminin kullanılmasıdır.

Özürlülük/Fonksiyonellik değerlendirmede BUT uygulaması ile ilgili literatür son beş yıl içinde yayınlanmaya başlamış olup; sayısı her geçen gün artmaktadır. Baş ağrısı, omuz problemleri, çocuklarda fiziksel fonksiyon, omurilik hasarlı hastalarda özürlülük, alt ekstremitte özürlülük durumu, depresif duygu durumu, osteoartritli hasta değerlendirimi, postakut rehabilitasyon hastalarında yapılmış BUT çalışmaları vardır. Bu çalışmaların bir kısmı gerçek BUT uygulaması olup (Ware ve ark., 2003; Hart ve ark., 2005; Haley ve ark., 2005; Hart ve ark., 2006;); bir kısmı benzetim çalışmasıdır (Theiler ve ark., 2002; Dijkers, 2003; Haley ve ark., 2004; Koestler ve ark., 2005). Genel olarak çalışmaların sonucunda; hastaların  $\theta$  düzeyi daha az sayıda madde ile daha doğru olarak tahmin edilmiş, zaman ve hasta bıkkınlığı açısından yöntemin önemli avantajlarının olduğu belirtilmiştir.

Yurtdışında iki sonuçlu maddeler kullanılarak BUT uygulamasını gerçekleştirebilen çeşitli ticari programlar bulunmaktadır. Ancak çok sonuçlu maddeler veya hem iki hem de çok sonuçlu maddelerden oluşan soru

bankaları ile BUT uygulaması yapan yazılım programı sadece proje bazlı geliştirilmiş olup; ticari kullanıma açık olan herhangi bir yazılım bulunmamaktadır. Bu çalışma kapsamında geliştirilen yazılım programı yardımıyla kalibrasyonu gerçekleştirilen iki ve çok sonuçlu maddeler için BUT uygulaması, “en çok olabilirlik” yöntemiyle yapılmıştır. Bu programın soru bankasındaki maddelere verdikleri yanıtları benzetim çalışması ile türetilen hasta grupları için BUT uygulaması yapabilen bir modülü de mevcuttur.

Benzetim çalışmaları, BUT yöntemi kullanılarak seçilen maddelere verilen yanıtların, gerçek BUT uygulamasında elde edilecek olanlarla aynı olduğu varsayımına dayalı olarak gerçekleştirilir. BUT yöntemine dayalı olarak yapılan çalışmalar (Dijkers, 2003; Ware, 2003), bu tür benzetim çalışmalarının gerçek BUT uygulamalarına oldukça yakın sonuçlar verdiğini göstermektedir. Bu tür benzetim çalışmaları tekrarlanarak,  $\theta$  düzeyinin tahmin edilmesinde kullanılabilecek en uygun parametre tahmin yöntemi, BUT yönteminin tamamlanması için optimal durdurma kriterleri, en iyi madde seçim yaklaşımı belirlenebilir.

Literatürde bilgisayar kullanılarak yapılan değerlendirmelerde hasta ya da bireylerin uyumu ile ilgili çalışmalar da vardır. Ware ve ark. (2003), Hart ve ark. (2005), Haley ve ark. (2005), Hart ve ark. (2006) tarafından yapılan BUT çalışmaları sonucunda hasta bıkkınlığında ve ölçeğin uygulama süresinde dikkate değer azalışlar olduğu belirtilmiştir. Bizim çalışmamızda da, BUT uygulaması sonucunda, hastaların özürülük düzeylerinin daha az sayıda madde (ortalama $\pm$ standart sapma:  $17\pm 3$ ) ile daha kısa sürede elde edildiği ( $6\pm 2$  dk.), soru bankasında yer alan tüm maddeler kullanılarak elde edilen sonuçlarla, BUT uygulaması sonucunda elde edilenler arasında yüksek ilişki olduğu belirlenmiştir. Ancak hastaların yaşlı, eğitim düzeylerinin düşük ve bilgisayar tecrübelerinin olmaması nedenleriyle BUT uygulamasını tek başlarına yapabilmeye zorluk yaşadıkları tespit edilmiştir. Bu tür sorunlarla karşılaşılacak hastalar için bir araştırma görevlisi yardımıyla uygulamayı tamamlamaları sağlanmıştır.

## **Çalışmanın Kısıtlılıkları**

Çalışma kapsamında karşılaşılan bazı kısıtlılıklar aşağıda verilmiştir:

1. Her bir hastalık grubu için soru bankalarının oluşturulması aşamasında yapılan klasik uygulamalar sonucunda bazı maddelere yanıt veren hasta sayısının oldukça az olduğu belirlenmiştir. Buna göre, bel ağırlı hastalar için geliştirilen soru bankası çalışmayan / cinsel olarak aktif olmayan; diz osteoartritli hastalar için geliştirilen soru bankası ise, çalışmayan / cinsel olarak aktif olmayan / küvette banyo yapmayan / ağır ev işi yapmayan hastalar için geçerlidir.
2. Geliştirilen soru bankalarının faktör yapısını doğrulamak için yapılan DFA analizlerinde aynı veri setleri yerine, aynı grup hastalığa sahip farklı hastalardan elde edilen veri setlerinin kullanılması gerekmektedir. Ancak, madde sayısı göz önünde bulundurulduğunda DFA analizinin farklı bir veri seti üzerinde gerçekleştirilmesi için en az 200 hasta gerekmektedir.
3. Benzetim çalışması ile türetilen yanıtlarla yapılan BUT uygulaması sonucunda Newton-Raphson algoritmasının yakınsamama, birden fazla kök bulma veya kök bulamama sorunları nedeniyle bazı hastaların özürülük düzeyleri elde edilememiştir. Bu nedenle ileride yapılacak olan çalışmalar için iteratif çözümlemelere gerek duymayan tahmin yöntemlerinin kullanılmasının gerekli olduğu düşünülmektedir.
4. Soru bankalarına, yeni maddeler ekleyerek  $\theta$  tüm düzeyleri için yeterli bilgi sağlayacak duruma getirmek gerekmektedir.
5. Soru bankalarından elde edilen sonuçlarla, benzer amaca hizmet eden ölçme araçlarından elde edilen sonuçların ilişkisine bakılmalıdır.

## 5. SONUÇ VE ÖNERİLER

Bel ağrısı ve diz osteoartriti sık görülen kronik kas-iskelet sistemi sorunları olup; bu hastalığa sahip bireylerin özürülük düzeylerinin objektif parametreler ve standart değerlendirme prosedürleri kullanılarak belirlenmesi önemlidir. Bu amaçla, çalışmamızda her bir hastalık grubu için mevcut ölçeklerde yer alan maddelerden ICF sınıflamasını temel alan soru bankaları oluşturulmuştur.

Literatürde, bel ağrılı ve diz osteoartritli hastaların özürülük durumunu değerlendiren bir soru bankası bulunmamakta, dolayısıyla özürülülüğü değerlendiren soru bankalarını oluşturacak maddelerin kaç boyutta toplanacağı ile ilgili de bir çalışma yoktur. Bu çalışmada soru bankalarındaki maddelerin kaç boyutta toplandığı kategorik verilerde AFA ve DFA ile incelenmiş; elde edilen modellerin yeterli uyuma sahip olduğu belirlenmiştir. Bu analizler sonucunda bel ağrılı hastalar için geliştirilen soru bankasında yer alan maddelerin “Ağrı-Uyku-Duygusal Durum” ve “Aktivite-Katılım” olarak, diz osteoartritli hastalar için geliştirilen soru bankasında yer alanların “Aktivite-Ağrı” ve “Ağrı-Sosyal-Bilişsel-Duygusal Durum” olacak şekilde iki boyutta toplandığı gözlenmiştir.

Soru bankalarının psikometrik özelliklerinin incelenmesinde Rasch türü modellerden Kısmi Kredi Modeli kullanılmış; modele uyum göstermeyen, tek boyutluluk varsayımını bozan ve MİF gösteren maddeler çıkartıldıktan sonra her bir hastalık grubu için elde edilen soru bankalarının hastaların özürülük değerlendiriminde güvenilir biçimde kullanılabilecekleri belirlenmiştir. Ancak, bu soru bankalarında  $\theta$ 'nın bazı düzeylerinde yeterli sayıda madde içermediği tespit edilmiş, bu düzeylere uygun yeni maddelerin eklenmesinin, soru bankalarının kalitesini arttıracığı sonucuna ulaşılmıştır.

Çalışma kapsamında, her bir hastalık grubu için elde edilen final soru bankalarının geçerliliğinin değerlendirilmesi için biri benzetim sonucunda elde edilen veriler, diğeri gerçek veriler üzerinde olmak üzere iki farklı uygulama gerçekleştirilmiştir. Bu amaçla, Rasch analizi sonucunda elde edilen final soru bankalarında yer alan maddeler kullanılarak BUT yönteminin uygulanabilmesi için bir yazılım programı geliştirilmiştir. Bu program temelde gerçek hastalar üzerinde BUT uygulaması yapabilecek şekilde tasarlanmış olup; benzetim çalışmaları için de bir modüle (CATSIM) sahiptir. Geliştirilen BUT programının sadece sağlık alanında değil eğitim ve sosyal bilimler alanlarında da kullanılabileceği; eğitim alanında, dünya çapında uygulanan TOEFL, GRE gibi testlerin benzerlerinin ülkemizde de uygulanabilmesine olanak tanıyacağı düşünülmektedir. Ayrıca, bu program ile verilerin analizi için bilgisayara veri girmeye ihtiyaç duyulmayacak, bu sayede farklı alanlarda çok merkezli çalışmaların standardizasyonu sağlanacak ve istenildiği takdirde çeşitli hastalıklar için ulusal düzeyde veri tabanları oluşturulabilecektir.

Her iki uygulama sonucunda da hastaların özürölülük düzeyleri daha az madde ile daha kısa sürede belirlenebilmiştir. Yapılan karşılaştırmalar sonucunda soru bankasında yer alan maddelerin tümü kullanıldığında elde edilen özürölülük düzeyleri ile BUT yöntemi sonucunda elde edilenler arasında yüksek uyum olduğu SKK, Spearman ilişki korelasyonu ve Blant-Altman grafiğı ile ortaya konmuştur. Herbir hastalık grubu için soru bankasındaki tüm maddeler sorulduğunda elde edilen özürölülük düzeyleri, BUT yöntemi uygulanarak ortalama  $17\pm3$  madde ile belirlenebilmektedir. Soru bankasında yer alan maddelerin tümünün uygulanması ortalama olarak  $15\pm2.5$  dakika sürerken; BUT uygulamasında bu sürenin yarı yarıya azaldığı belirlenmiştir ( $6\pm2$  dk). Hastalara BUT yöntemi ile ilgili memnuniyet durumları sorulduğunda, yaş ortalamaları yüksek, eğitim düzeyleri düşük ve bilgisayar tecrübeleri olmadığından BUT yönteminin daha zor olduğunu belirtmişlerdir.

Bu alıřma sonucunda, her bir hastalık grubu iin geerlilik ve gvenilirlikleri modern yntemlerle kanıtlanan soru bankalarında yer alan maddeler kullanılarak, bel ađrılı ve diz osteoartritli hastaların zrllk deđerlendirmenin daha kısa srede, daha az madde ile yeterli kesinlikte elde edilebildiđi gsterilerek hedeflenen amaların tmne bařarıyla ulařılmıřtır.

## ÖZET

### **Kas-İskelet Sistemi Sorunlarının Özürlülük Değerlendirmesinde Bilgisayar Uyarlamalı Test Yönteminin Uygulanması**

Kas-iskelet sistemi hastalıkları, kronik, özürlülüğe yol açan sorunlardır. Bu hastalıkların, gerek yüksek prevalansları gerekse sağlık hizmetlerindeki yüksek maliyetleri nedeniyle toplumdaki ekonomik yükü de fazladır. Kas-iskelet sistemi sorunlarının tedavi ve rehabilitasyonunda hedef, hastalık aktivitesini kontrol altına almak, fonksiyonları korumak ve geliştirmek ve sonuçta özürlülüğü engellemektir. Özürlülük değerlendirmenin etkin bir şekilde yapılabilmesi, ancak geçerli, güvenilir, değişime duyarlı ve kolay uygulanabilir değerlendirme araçları ile mümkündür. BUT, kalibrasyonu yapılmış soru bankalarından sadece en fazla bilgiyi sağlayan maddeleri seçerek, ölçüm kesinliğinin artmasını ve hastaların bıkkınlık düzeylerinin azalmasını sağlar.

Bu çalışmanın amacı, toplumda sık görülen ve özürlülüğe yol açan kas-iskelet sistemi hastalıklarından bel ağrısı ve diz osteoartritinde BUT yöntemiyle hastaların özürlülük düzeylerini belirlemektir. Bu amaca yönelik olarak öncelikle ICF sınıflaması temel alınarak bel ağrısı ve diz osteoartritinde özürlülük değerlendirimi için birer soru bankası oluşturulmuş; 266 bel ağrısı ve 225 diz osteoartriti hastası, sırasıyla 108 ve 110 maddeye yanıt vermiştir. Soru bankalarının boyut yapısı kategorik veriler için AFA ve DFA analizleri ile psikometrik özellikleri kısmi kredi modeli ile değerlendirilmiştir. İkinci aşamada da final soru bankaları kullanılarak, geliştirilen yazılımla sağlık alanında yeni bir uygulama olan BUT yöntemiyle "en çok olabilirlik yöntemi" kullanılarak hastaların özürlülük düzeyi belirlenmiştir.

Elde edilen sonuçlar, soru bankasındaki tüm maddeler ile elde edilen özürlülük düzeylerinin, BUT yöntemi kullanıldığında daha az madde ile belirlenebildiğini ve elde edilen sonuçlar arasında yüksek derecede uyum olduğunu göstermiştir. Bu sonuçlar, özürlülüğün BUT yöntemi kullanılarak doğru ve klinik olarak etkin bir şekilde değerlendirilebileceğini kanıtlamaktadır.

**Anahtar Sözcükler:** Bilgisayar uyarlamalı test, ölçek, Rasch modelleri, rehabilitasyon, sonuç ölçümü.

## SUMMARY

### **An Application of Computerized Adaptive Testing in the Evaluation of Disability in Musculoskeletal Disorders**

Musculoskeletal diseases are chronic and disabling disorders. The economic burden of these disorders on the society is great due to both high prevalence and high cost of the diseases to the health care system. The main goals in the treatment and rehabilitation of musculoskeletal disorders are to control the disease activity, maintain and improve function and consequently prevent disability. Assessment of disability in an effective manner will only be possible with valid, reliable, responsive and easily applicable assessment instruments. Computerized Adaptive Tests (CAT) promise to enhance measurement precision and reduce respondent's burden by only selecting the most informative items from a calibrated item bank.

The purpose of this study is to determine the disability levels of patients with low back pain and knee osteoarthritis, common and disabling musculoskeletal diseases, by using computerized adaptive testing (CAT) method. To achieve this purpose, firstly, two item banks each for the assessment of disability in low back pain and knee osteoarthritis have been developed based on the ICF model. Two hundred sixty six patients with low back pain and two hundred twenty five patients with knee osteoarthritis have answered 108 and 110 questions respectively. The dimensionality of item banks has been analyzed using exploratory and confirmatory factor analyses for categorical data and the psychometric properties of them have been evaluated by Partial Credit Model. At the second stage, CAT based disability levels of patients have been determined with "maximum likelihood method" from these final item banks by using the CAT software that had been developed in this study.

Results indicate that comparable disability levels to the full-length item banks can be achieved in the CAT version using fewer items and there is a high agreement between them. These results support further consideration of using CAT programs for accurate and effective clinical assessments of disability.

**Key Words:** Computerized adaptive test, outcome measurement, Rasch models, rehabilitation, scale.



## KAYNAKLAR

- ALPAR, R. (2003). Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş 1. Ankara: Nobel Yayın Dağıtım.
- ANIL, D. (2002). Deneme Uygulamasının Yapılamadığı Durumlarda Madde ve Test Parametrelerinin Klasik ve Örtük Özellikler Test Teorilerine Göre Kestirilmesi. Doktora Tezi, Hacettepe Üniv. Sosyal Bilimler Enstitüsü.
- BAKER, F.B. (2001). The Basics of Item Response Theory. USA: ERIC Clearinghouse on Assessment and Evaluation.
- BARNETT, V.D. (1966). Evaluation of the maximum-likelihood estimator where the likelihood equation has multiple roots. *Biometrika*, **53** (1): 151-65.
- BEATON, D.E., BOMBARDIER, C., GUILLEMIN, F., FERRAZ, M.B. (2000). Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures. *Spine*, **25** (24): 3186–91.
- BJORNER, J.B., KOSINSKI, M., WARE, J.E. (2003). Calibration of an Item Pool for Assessing the Burden of Headaches: An Application of Item Response Theory to the Headache Impact Test (HIT™). *Quality of Life Research*, **12**: 913-933.
- BLAIS, J.G., RAICHE, G. (2002). Some Features of the Sampling Distribution of The Ability Estimate in Computerized Adaptive Testing According to Two Stopping Rules. Erişim: [[http://eric.ed.gov/ERICDocs/data/ericdocs2/content\\_storage\\_01/0000000b/80/0d/e1/36.pdf](http://eric.ed.gov/ERICDocs/data/ericdocs2/content_storage_01/0000000b/80/0d/e1/36.pdf)]. Erişim Tarihi: 24.04.2006
- BLAND, J.M., ALTMAN, D.G. (1995). Multiple Significance Tests: The Bonferroni Method. *BMJ*, **310**:170.
- CHANG, C., REEVE, B.B. (2005). Item Response Theory and Its Applications To Patient-Reported Outcomes Measurement. *Evaluation&The Health Professions*, **28** (3): 264-282.
- CIEZA, A., BROCKOW, T., EWERT, T., AMMAN, E., KOLLERITS, B., CHATTERJI, S., ÜSTÜN, T.B., STUCKI, G. (2002). Linking Health-Status Measurements to the International Classification of the Functioning Disability and Health. *J Rehabil Med*, **34**: 205-210.
- CIEZA, A., STUCKI, G., WEIGL, M., DISLER, P., JÄCKEL, W., VAN DER LINDEN, S., KOSTANJSEK, N., DE BIE, R. (2004). ICF Core Sets For Low Back Pain. *Journal of Rehabilitation Medicine*, **36** (S44): 69-74.

- DAVIDSON M., KEATING J.L., SOPHIE M.A. (2004). A Low Back-Specific Version of the SF-36 Physical Functioning Scale. *Spine*, **29 (5)**: 586-594.
- DIJKERS, M.P. (2003). A Computer Adaptive Testing Simulation Applied to the FIM Instrument Motor Component. *Arch Phys Med Rehabil*, **84**: 384-93.
- DOĞAN, N. (2002). Klasik Test Kuramı ve Örtük Özellikler Kuramının Örneklemeler Bağlamında Karşılaştırılması. Doktora Tezi, Hacettepe Üniv. Sosyal Bilimler Enstitüsü.
- DOWNING, S.M. (2003). Item Response Theory: Applications of Modern Test Theory in Medical Education. *Medical Education*, **37**: 739-745.
- DREINHOFER, K., STUCKI, G., EWERT, T., HUBER, E., EBENBICHLER, G., GUTENBRUNNER, C., KOSTANJSEK, N., CIEZA, A. (2004). ICF Core Sets for Osteoarthritis. *Journal of Rehabilitation Medicine*, **36 (S44)**: 75-80.
- EMBRETSON, S.E., REISE, S.P. (2000). Item Response Theory for Psychologists. Mahwah, NJ: Erlbaum.
- FLIEGE, H., BECKER, J., WALTER, O.B., BJORNER, J.B., KLAPP, B.F. ROSE, M. (2005). Development of a CAT for Depression (D-CAT). *Qual Life Res*, **14 (10)**:2277-91.
- FUHRER, M.J. (1994). Subjective Well-Being: Implications for Medical Rehabilitation Outcomes and Models of Disablement. *Am J Phys Med Rehabil.*, **73**: 358-64.
- GOEGEBEUR, Y., BOECK, P., MOLENBERGHS, G., PINO, G. (2006). A Local Influence Based Diagnostic Approach to A Speeded IRT Model. Erişim: [[http://ucs.kuleuven.be/seminars\\_events/files/abstractGoegebeur.pdf](http://ucs.kuleuven.be/seminars_events/files/abstractGoegebeur.pdf)]. Erişim Tarihi: 19.06.2006
- GORIN, J.S., DODD, B.G., FITZPATRICK, S.J., SHIEH, Y.Y. (2005). Computerized Adaptive Testing with the Partial Credit Model: Estimation Procedures, Population Distributions, and Item Pool Characteristics. *Applied Psychological Measurement*, **29 (6)**: 433-456.
- GREEN, B.F., BOCK, R.D., HUMPHREYS, L.G., LINN, R.L., RECKASE, M.D. (1984). Technical Guidelines for Assessing Computerized Adaptive Tests. *Journal of Educational Measurement*, **21 (4)**: 347-60.
- GUILLEMIN, F. (2000). Functional Disability and Quality of Life Assessment in Clinical Practice. *Rheumatology*, **39 (Suppl1)**: 17-23.

- HALEY, S.M., RACZEK, A.E., COSTER, W.J., DUMAS, H.M., FRAGALA-PINKHAM, M.A. (2005). Assessing Mobility in Children Using a Computer Adaptive Testing Version of The Pediatric Evaluation of Disability Inventory. *Arch. Phys Med Rehabil.*, 86 (5): 932-9.
- HALEY, S.M., COSTER, W.J., ANDRES, P.L., KOSINSKI, M.(2004). Score Comparability of Short Forms and Computerized Adaptive Testing: Simulation Study with the Activity Measure for Post-Acute Care. *Arch.Phys.Med.Rehabil.*, **85**: 661-666.
- HAMBLETON, R.K., SWAMINATHAN, H. (1985). Item Response Theory: Principles and Applications. Boston: Kluwer-Nijhoff Publishing.
- HAMBLETON, R.K., SWAMINATHAN, H., ROGERS, H.J. (1991). Fundamentals of Item Response Theory. California: SAGE Publications.
- HART, D.L, COOK, K.F., MIODUSKI, J.E., TEAL, C.R., CRANE, P.K. (2006). Simulated computerized adaptive test for patients with shoulder impairments was efficient and produced valid measures of function. *Journal of Clinical Epidemiology*, **59**: 290-298.
- HART, D.L., MIODUSKI, J.E., STRATFORD, P.W. (2005). Simulated computerized adaptive tests for measuring functional status were efficient with good discriminant validity in patients with hip, knee, or foot/ankle impairments. *Journal of Clinical Epidemiology*, **58**: 629-638.
- HAYS, R.D., MORALES, L.S., REISE, S.P. (2000). Item Response Theory and Health Outcomes Measurement in the 21<sup>st</sup> Century. *Medical Care*, **38 (9 suppl.II)**: 28-42.
- HULIN, C. L., DRASGOW, F., PARSONS, C.K. (1983). Item Response Theory: Application to Psychological Measurement. Homewood, IL: Dow-Jones, Irwin.
- KAZIZ, L., ANDERSON, J.J., MEENAN, R.F. (1989). Effect Sizes for Interpreting Changes in Health Status. *Med Care*, **27**: 178-189.
- KINGSBURY, G., ZARA, A. (1989). Procedures for Selecting Items for Computerized Adaptive Tests. *Applied Measurement in Education*, **2(4)**, 359-75.
- KOESTLER, M.E., LIBBY, E., SCHOFFERMAN, J., REDMOND, T. (2005). Web-Based Touch-Screen Computer Assessment of Chronic Low Back Pain. *CIN: Computers, Informatics, Nursing*, **23 (5)**: 275-284.

- KOPEC, J.A., SAYRE, E.C., DAVIS, A.M., BADLEY, E.M., ABRAHAMOWICZ, M., SHERLOCK, L., WILLIAMS, J.I., ANIS, A.H., ESDAILE, J.M.: Assessment of health-related quality of life in arthritis: conceptualization and development of five item banks using item response theory. *Health and Quality of Life Outcomes*, **4**: 33, 2006.
- KÜÇÜKDEVECİ, A. (2004). Romatizmal Hastalıkların Değerlendiriminde Kullanılan Ölçekler, I. Ulusal Romatizmal Hastalıklar Kongresi, 28 Nisan-2 Mayıs 2004, Antalya, Türkiye.
- LINACRE, J.M. (2000). Computer-Adaptive Testing: A Methodology Whose Time Has Come. Erişim: [<http://www.rasch.org/memo69.pdf>]. Erişim Tarihi: 21.06.2004
- LORD, F.M. (1980). Applications of Item Response Theory to Practical Testing Problems. Hillsdale, New Jersey: Lawrence Erlbaum Associates Inc.
- MARTIN, M., KOSINSKI, M., BJORNER, J.B., WARE, J.E. Jr, MACLEAN, R., LI, T. (2007). Item Response Theory Methods Can Improve the Measurement of Physical Function by Combining The Modified Health Assessment Questionnaire and The SF-36 Physical Function Scale. *Quality of Life Research*, **16**: 647–660.
- McHORNEY, C.A., MONAHAN, P.O. (2004). Postscript: Applications of Rasch Analysis in Health Care. *Med. Care*, **42(1 Suppl)**: 173-8.
- MATHEWS, J.H., FINK, K.D. (2004). Numerical Methods Using Matlab (Fourth Edition (& Higher)). Upper Saddle River: Pearson Prentice Hall.
- MICROSOFT INC. (2006). History of Adaptive Testing in High Stakes Testing. Erişim: [[http://www.windowsgalore.com/cert/adaptive\\_testing/history.html](http://www.windowsgalore.com/cert/adaptive_testing/history.html)]. Erişim Tarihi: 25.12.2006.
- MUTHEN&MUTHEN (2006). MPlus, Version 4.1. Los Angeles, CA.
- OWEN, R.J. (1975). A Bayesian Sequential Procedure for Quantal Response in The Context of Adaptive Mental Testing. *Journal of American Statistical Association*, **70**: 351-356.
- ÖZTUNA, D., ELHAN, A.H., TÜCCAR, E. (2006). The Investigation of Four Different Normality Tests in Terms of Type I Error Rate and Power Under Different Distributions. *Turkish Journal of Medical Sciences*, **36**: 171-176.
- PALLANT, J.F., TENNANT, A. (2007). An Introduction to the Rasch Measurement Model: An Example Using the Hospital Anxiety and Depression Scale (HADS). *The British Journal of Clinical Psychology*, **46 (Pt 1)**: 1–18.

- RASCH, G. (1980). Probabilistic models for some intelligence and attainment tests. Chicago, IL: MESA Press.
- REEVE, B.B. (2002). An Introduction to Modern Measurement Theory. Eriřim: [<http://appliedresearch.cancer.gov/areas/cognitive/immt.pdf>]. Eriřim Tarihi: 13.02.2004
- RUDNER, L.M. (1998). An On-Line, Interactive, Computer Adaptive Testing Tutorial. Eriřim: [<http://edres.org/scripts/cat>]. Eriřim Tarihi: 12.01.2006.
- RUMM LABORATORY PTY LTD. (2007). RUMM2020, Version 4.1. Duncraig, WA: RUMM Laboratory Pty Ltd.
- SCANTRON CORP. (2003). Fundamental Issues Regarding Computer-Adaptive Testing. Eriřim: [[www.scantron.com/downloads/pdf/CAT-Paper.pdf](http://www.scantron.com/downloads/pdf/CAT-Paper.pdf)]. Eriřim Tarihi: 21.06.2004.
- SEGALL, D.O. (2003). Computerized Adaptive Testing. Encyclopedia of Social Measurement. Eriřim: [[www.academicpress.com/refer/measure/measarts.htm](http://www.academicpress.com/refer/measure/measarts.htm)]. Eriřim Tarihi: 02.01.2008.
- SUAREZ-ALMAZOR, M.E., KENDALL, C., JOHNSON, J.A., SKEITH, K., VINCENT, D. (2000). Use of Health Status Measures in Patients with Low-Back Pain in Clinical Settings. Comparison of Specific, Generic and Preference-Based Instruments. *Rheumatology*, **39**: 783-790.
- SUHR, D.D. (2006). Exploratory or Confirmatory Factor Analysis? Eriřim: [[www2.sas.com/proceedings/sugi31/200-31](http://www2.sas.com/proceedings/sugi31/200-31)]. Eriřim Tarihi: 24.12.2007.
- ŐİMŐEK, Ő.F. (2007). Yapısal Eřitlik Modellemesine Giriř: Temel İlkeler ve LISREL Uygulamaları. İstanbul: Ekinoks Yayıncılık.
- TEKİN, H. (2002). Eđitimde Őlçme ve Deđerlendirme. Ankara: Yargı Yayıncılık.
- TENNANT, A., CONAGHAN, P.G. (2007). The Rasch Measurement Model in Rheumatology: What Is It and Why Use It? When Should It Be Applied, and What Should One Look for in a Rasch Paper? *Arthritis&Rheumatism*, 57 (8): 1358-1362.
- TENNANT, A., PENTA, M., TESIO, L., GRIMBY, G., THONNARD, J.L., SLADE, A., SIMONE, A., CARTER, J. [LUNDGREN-NILSSON, A.](#), [TRIPOLSKI, M.](#), [RING, H.](#), [BIERING-SŐRENSEN, F.](#), [MARINCEK, C.](#), [BURGER, H.](#), [PHILLIPS, S.](#) (2004). Assessing and Adjusting for Cross-Cultural Validity of Impairment and Activity Limitation Scales Through Differential Item Functioning within the Framework of the Rasch Model. the PRO-ESOR project. *Med Care*, **42 (Suppl1)**: 37-48.

- THEILER, R., SPIELBERGER, J., BISCHOFF, H.A., BELLAMY, N., HUBER, J., KROESEN, S. (2002). Clinical Evaluation of the WOMAC 3.0 OA Index in Numerical Rating Scale Format Using a Computerized Touch Screen Version. *Osteoarthritis and Cartilage*, **10**: 479-481.
- THISSEN, D., ORLANDO, M. (2001). Item Response Theory for Items Scored in Two Categories. Hillsdale, NJ: Erlbaum, Chapter 3.
- THISSEN, D., STEINBERG, L., & WAINER, H. (1988). Use of Item Response Theory in the Study of Group Differences in Trace Lines. NJ: LEA.
- TIKU, ML. (1967). Estimating Mean and Standard Deviation from Censored normal sample. *Biometrika*, **54**: 155.
- TURGUT, M.F. (1977). Eğitimde Ölçme ve Değerlendirme Metodları. Ankara: Nüve Matbaası.
- VAUGHAN, D.C. (1992) On the Tiku-Suresh Method of Estimation. *Communications in Statistics-Theory and Methods*, **21**: 451.
- WADE, D.T. (2003). Outcome Measures for Clinical Rehabilitation Trials. *Am J Phys Med Rehabil*, **82 (Suppl)**: 26-31.
- WAINER, H., DORANS, N.J., FLAUGHTER, R., GREEN, B.F., MISLEVY, R.J., STEINBERG, L., THISSEN, D. (2000). Computerized Adaptive Testing: A Primer. New Jersey: Lawrence Erlbaum Associates, Inc.
- WARE, J.E. (2003). Conceptualization and Measurement of Health-Related Quality of Life: Comments on an Evolving Field. *Arch.Phys.Med.Rehabil.*, **84 (Suppl. 2)**: S43–51.
- WARE, J.E., GANDEK, B., SINCLAIR, S.J., BJORNER, J.B. (2005). Item Response Theory and Computerized Adaptive Testing: Implications for Outcomes Measurement in Rehabilitation. *Rehabilitation Psychology*, **50 (1)**, 71–78.
- WARE, J.E., KOSINSKI, M., BJORNER, J.B., BAYLISS, M.S., BATENHORST, A., DAHLÖF, C.G.H.,TEPPER, S., DOWSON, A. (2003). Applications of Computerized Adaptive Testing (CAT) to the Assessment of Headache Impact. *Quality of Life Research*, **12**: 935-952.
- WEISS, D.J. (1982). Improving Measurement Quality and Efficiency with Adaptive Testing. *Applied Psychological Measurement*, **6**: 473-92.
- WHALLEY, D., McKENNA, S.P., DE JONG, Z., VAN DER HEIJDE, D. (1997). Quality of Life in Rheumatoid Arthritis. *Br J Rheumatol.*, **36**:884-8.

WHO (2001). International Classification of Functioning, Disability, and Health. Geneva, Switzerland: World Health Organization.

WIBERG, M. (2004). Classical Test Theory vs. Item Response Theory an Evaluation of the Theory Test in the Swedish Driving-License Theory. Eriřim: [<http://www.umu.se/edmeas/publikationer/pdf/EM%20no%2050.pdf>]. Eriřim Tarihi: 23.02.2005.

WRIGHT, B.D., MASTERS, G.N. (1982). Rating Scale Analysis. Chicago: Mesa Press p: 43.

WRIGHT, B.D., STONE, M.H. (1979). Best Test Design Rasch Measurement. Chicago: Mesa Pres, p: 21.

# ÖZGEÇMİŞ

## I- Bireysel Bilgiler

**Adı ve Soyadı:** Derya ÖZTUNA

**Doğum Yeri ve Tarihi:** ANKARA-19.04.1978

**Uyuşu:** T.C.

**Medeni Durumu:** Evli

**İletişim Adresi ve Telefonu:** Ankara Üniversitesi Tıp Fakültesi Biyoistatistik ABD.

Morfoloji Kampüsü 06700 Sıhhiye/ANKARA

Tel: (0312) 3103010/375

## II- Eğitim

<b>Doktora</b>	Ankara Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı Mezuniyet yılı: 2008
<b>Yüksek Lisans</b>	Hacettepe Üniversitesi Fen-Edebiyat Fakültesi İstatistik Bölümü Mezuniyet yılı: 2002
<b>Lisans</b>	Hacettepe Üniversitesi Fen-Edebiyat Fakültesi İstatistik Bölümü Mezuniyet yılı: 1999
<b>Lise</b>	Kurtuluş Lisesi Mezuniyet Yılı: 1995

**Yabancı Dili:** İngilizce

## III- Ünvanları

2003 – 2008: Araştırma Görevlisi; Ankara Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı

## IV- Mesleki Deneyimi

**ANKARA  
ÜNİVERSİTESİ TIP  
FAKÜLTESİ  
BİYOİSTATİSTİK  
ANABİLİM DALI  
ANKARA**

**Pozisyon:** 2003 yılından itibaren "Araştırma Görevlisi" olarak çalışmaktayım.  
**Özet İş Tanımı:** Tıp Fakültesi 1., 2. ve 3. sınıf öğrencilerine Bilgisayar ve Biyoistatistik derslerinin verilmesi, fakülte kapsamında ya da diğer üniversite ve kurumlardan gelen araştırmacıların çalışmalarının istatistiksel analizlerinin yapılması, istatistiksel danışmanlık verilmesi

**TED ANKARA  
KOLEJİ  
ANKARA**

**Pozisyon:** 01.07.2000 – 01.06.2002 tarihleri arasında İstatistikçi olarak görev aldım.  
**Özet İş Tanımı:** Öğrencilerin okul içi ve okul dışı (ÖSS, LGS, MEB Ortaöğretim Kurumları sınavı v.s.) sınav sonuçlarının değerlendirilmesi, ölçme-değerlendirme çalışmalarının yapılması, istatistiksel analizlerin yapılması ve anket düzenlenmesi ile değerlendirilmesi.



## **V- Üye Olduğu Bilimsel Kuruluşlar**

İstatistik Mezunları Derneği

International Biometrics Society (Eastern Mediterranean Region)

Hacettepe Mezunlar Derneği

## **VI- Bilimsel İlgi Alanları**

### **YAYINLARIN LİSTESİ**

1. Yanık, B., Atalar, H., Külcü, D., Öztuna, D.: The Effect Of Body Mass Index On Bone Mineral Density In Postmenopausal Women - Postmenopozal Kadınlarda Vücut Kütle İndeksinin Kemik Mineral Yoğunluğuna Etkisi. Osteoporoz Dünyasından, 13:56-59, 2007.
2. Ulusoy, M.G., Kankaya, Y., Uysal, A., Sungur, N., Koçer, U. Kankaya, D., Öztuna, D.: "Lid Technique":Cyanoacrylate Assisted Anastomosis Of Small Sized Vessels. J Plast Recons Aesthet Surg (Kabul edildi)
3. Akay, G.G., Özkal, P., Rüstemoğlu, A., Varol, N., Akçora, D., Oskay, S., Altınok, B., Karadağ, A., Özkan, T., Pekurnaz, G., Öztuna, D., Sunguroğlu, A.: Imatinib Causes Fragile Site Expression in CML Patients and Chromosome Breakage in Cultured Human Peripheral Blood Lymphocytes. Exp Oncol (Yayın aşamasında)
4. Kaymak Ç. Karahalil B., Özcan N.N., Öztuna D.: Association Between GSTP1 Gene Polymorphism and Serum Alpha-GST Concentrations Undergoing Sevoflurane Anaesthesia. European Journal of Anaesthesiology; : 1-7 (Yayın aşamasında)
5. Öztuna D., Kurşun N., Elhan A.H.: Sağlık Araştırmalarında Kullanılan İlişki Katsayıları. Türkiye Klinikleri Dergisi. (Yayın aşamasında)
6. Öztuna D., Gökçe E.: Yükseköğretimde Öğrenci Başarısının Artırılmasında İç ve Dış Motivasyon Faktörlerinin Etkisi. Çağdaş Eğitim Dergisi, 335: 37-46, 2006.
7. Kaymak Ç., Kocabaş N.A, Durmaz E., Öztuna D.:  $\beta 2$  Adrenoceptor (ADRB2) Pharmacogenetics and Cardiovascular Phenotypes During Laryngoscopy and Tracheal Intubation. Int J Toxicol., 25 (6): 443-9, 2006.
8. Öztuna D., Elhan A.H., Tüccar E.: The Investigation of Four Different Normality Tests in Terms of Type I Error Rate and Power Under Different Distributions. Turkish Journal of Medical Sciences, 36: 171-176, 2006.
9. Özçakar Z.B., Yalçınkaya F., Yüksel S., Acar B., Gökmen D., Ekim M.: Possible Effects of Subclinical Inflammation on Daily Life in Familial Mediterranean Fever, Clin Rheumatol 2006 Mar;25(2):149-52.
10. Genç Y., Gökmen D., Tüccar E., Yağmurlu B.: Estimation of Sensitivity and Specificity for Clustered Data. Turkish Journal of Medical Sciences, 35 (1): 21-24, 2005.
11. Togay-Işıkkay C., Kim J., Betterman K., Andrews C., Meads D., Tesh P., Tegeler C., Öztuna D.: Carotid Artery Tortuosity, Kinking, Coiling: Stroke Risk Factor, Marker, Or Curiosity?. Acta Neurol Belg, 105, 68-72, 2005.
12. Gökmen D., Genç Y., Atakurt Y., Yağmurlu B.: Kümelenmiş Verilerde Bağımlı İki Oranın Karşılaştırılması. Journal of Statistical Research, 3 (3): 21-29, 2004.
13. Gökmen D., Toktamış O.: Bant Genişliği Seçiminde Kullanılan Yöntemlerin Simetrik ve Simetrik Olmayan Dağılımlarda Karşılaştırılması. Journal of Statistical Research, 2(2): 17-30, 2003.

## ULUSLARARASI KONGRELERDE SUNULAN BİLDİRİLER ve POSTERLER

1. Avcı A., Namuslu M., Öztuna D., Ergüder İ.B., Devrim E., Sözener U., Kocaoğlu H., Durak İ.: Effects of Aqueous Red Clover Extract on DNA Turn-Over Enzymes in Cancerous and Non-Cancerous Human Gastric and Colon Tissues. The 11th International Congress Phytopharm 2007, 27-30 June 2007, Leiden, The Netherlands (Poster Presentation)
2. Candemir, B., Ertaş, F., Özdol, C., Atmaca, Y., Akyürek, O., Öztuna, D., Kumbasar, D., Erol, C.: Effect of Diabetes Mellitus on Coronary Arterial Remodeling Patterns in Patients with Minimal Atherosclerosis: An Ivus Study. 3th Clinical Vaskular Biology Congress, 25-29 April 2007, Antalya, Turkey. European Journal of Clinical Investigation, 37: 320-321, 2007 (Poster Presentation).
3. Kocabaş, N.A., Kaymak, Ç., Aydın, N., Öztuna, D., Karakaya, A.E.: Investigation of the  $\beta_2$ -Adrenoceptor (ADRB2) 16 and Glutathione S-transferase P1 (GSTP1) Gene Polymorphisms in Turkish Asthma Patients. The 44th Congress of the European Societies of Toxicology, 7-10 October, 2007, Amsterdam, The Netherlands. Toxicology Letters, 172: 164-165, 2007 (Poster Presentation).
4. Candemir, B., Ertaş, F.S., Özdol, C., Atmaca, Y., Öztuna, D., Akyürek, O., Kumbasar, D., Erol, C.: ECoronary arterial remodeling patterns in patients with or without hypertension determined by intravascular ultrasound. World Congress of Cardiology 2006, 2-6 September 2006, Barcelona, Spain. European Heart Journal, 27: 374, 2006 (Poster Presentation).
5. Sunguroğlu A., Akay GG., Özkal P., Varol N., Akcora D., Altinok B., Gökmen D., Avcı A., Ergüder İ.B., Devrim E., Durak İ.: Antiproliferative and Apoptotic Effects of Garlic on Chronic Myeloid Leukemia Cell Line. 54<sup>th</sup> Annual Congress on Medicinal Plant Research, 29 August - 2 September 2006, Helsinki, Finland. Planta Medica, 72: 1057, 2006 (Poster Presentation).
6. Durak İ., Sunguroğlu A., Avcı A., Devrim E., Ergüder İ.B., Akay GG., Özkal P., Varol N., Akcora D., Altinok B., Gökmen D.: Effects of Aqueous Garlic Extract on Oxidant/Antioxidant Status in 32D and 32 Dp Cell Lines. 54<sup>th</sup> Annual Congress on Medicinal Plant Research, 29 August - 2 September 2006, Helsinki, Finland. Planta Medica, 72: 1058, 2006 (Poster Presentation).
7. Öztuna D., Genç Y., Yüksel S., Ekim M.: Statistical Methods for Paired Clustered Binary Data. Abstract Book of 27th Annual Conference of the International Society for Clinical Biostatistics, p: 179, Geneva, 27-31 August 2006 (Poster Presentation-Conference Award for Scientists Winner).
8. Güngör, E., Dalya, K., Akçağlayan, E.S., Öztuna, D., Beksaç, M.: Role of Killer Inhibitory Receptor (KIR) Allele and Ligand Mismatches (mm) in HLA Identical Sibling Haematopoietic Stem Cell Transplants. Abstract Book of 20th European Immunogenetics and Histocompatibility Conference, 8-11 June 2006, Oslo, Norway. Tissue Antigens, 67 (6): 502 P112, 2006.
9. Özçakar B., Yalçinkaya F., Yüksel S., Acar B., Gökmen D., Ekim M.: The Possible Effect of Subclinical Inflammation on Daily Life in Familial Mediterranean Fever. Familial Mediterranean Fever (FMF) and Beyond. The Fourth International Congress on Systemic Autoinflammatory Diseases, Bethesda, 6-10 November 2005 (Poster Presentation).
10. İleri (Akçaboy) M., Özçakar B., Acar B., Yüksel S., Gökmen D., Ekim M., Yalçinkaya F.: Antihyperlipidemic Therapy in Children with Refractory Nephrotic Syndrome. 39. Annual Meeting of the European Society for Pediatric Nephrology, İstanbul, 10-23 September 2005 (Poster Presentation).

11. Genç Y, Öztuna D.: Closed Form Methods to Compare Two Proportions for Clustered Data. Abstract Book of Annual Conference International Society for Clinical Biostatistics, p: 146, Hungary, 21-25 August 2005 (Poster Presentation).

## **ULUSAL KONGRELERDE SUNULAN BİLDİRİLER ve POSTERLER**

1. Yanık, B., Akgedik, R., Özol, D., Öztuna, D.: Postmenopozal Dönemdeki Astımlı Hastalarda Obezite, İnhaler Steroid ve Osteoporoz İlişkisi. 29. Ulusal Solunum Konresi, 28 Ekim-31 Ekim 2007 (Poster Sunumu).
2. Kutlay, Ş., Küçükdeveci, A., Elhan, A., Öztuna, D., Koç, N., Gönül, G.: WHO-DAS-II Disabilite Değerlendirme Ölçeğinin Osteoartritli Hastalarda Geçerlilik ve Güvenilirliğinin Araştırılması. 21. Ulusal Fiziksel Tıp ve Rehabilitasyon Kongresi, Antalya, 24-29 Ekim 2007 (Poster Sunumu).
3. Öztuna, D., Küçükdeveci, A., Kutlay, Ş., Elhan, A.: WHO-DAS-II Disabilite Değerlendirme Ölçeğinin Bel Ağrılı Hastalarda Geçerlilik ve Güvenilirliğinin Araştırılması. 21. Ulusal Fiziksel Tıp ve Rehabilitasyon Kongresi, Antalya, 24-29 Ekim 2007 (Poster Sunumu).
4. Akay, G.G., Özkal, P., Rüstemoğlu, A., Varol, N., Akçora, D., Oskay, S., Altınok, B., Karadağ, A., Özkan, T., Öztuna, D., Sunguroğlu, A.: İmanitib (Gleevec), Kültüre Edilmiş Periferik Kan Lenfositlerinde Kromozom Kırığına İndüklemektedir. X. Ulusal Tıbbi Biyoloji ve Genetik Kongresi Bildiri Kitabı, sayfa: 238, 6-9 Eylül 2007 (Poster Sunumu)
5. Taşpınar M., Düzen, V., Özkal, P., Akpınar, P., Öztuna, D., Tutar, E., Sunguroğlu, A.: Myeloperoksidaz (MPO) 129G/A Poliformizmi ve Ateroskleroz. X. Ulusal Tıbbi Biyoloji ve Genetik Kongresi Bildiri Kitabı, sayfa: 173, 6-9 Eylül 2007 (Poster Sunumu)
6. Taşpınar M., Aydos, S., Düzen, V., Öztuna, D., Tutar, E., Sunguroğlu, A.: CYP1A1 Poliformizmi ve Ateroskleroz Riski. X. Ulusal Tıbbi Biyoloji ve Genetik Kongresi Bildiri Kitabı, sayfa: 191, 6-9 Eylül 2007 (Poster Sunumu)
7. Genç, Y., Ateş, C., Öztuna, D., Gültekin, S., Tüccar, E., Akman, M.: Kümelenmiş Verilerde İşlem Karakteristiği Eğrisi (İKE) Altında Kalan Alanın Tahmini. X. Ulusal Biyoistatistik Kongresi Bildiri Özetleri Kitabı, sayfa: 34, 5-8 Eylül 2007 (Sözlü Bildiri)
8. Genç, Y., Öztuna, D., Akman, M.: Kümelenmiş Verilerde Oranların Karşılaştırılması: İteratif Olmayan Yöntemleri İçeren Programın Tanıtılması. 5. İstatistik Kongresi ve Risk Ölçümleri ve Yükümlülük Toplantısı Bildiri Özetleri Kitabı, sayfa: 134-135, 20-24 Mayıs 2007 (Sözlü Bildiri)
9. Şahin, M., Özmen, S., Arslan, Z., Fırat, S., Öztuna, D.: Dr. Sami Ulus Çocuk Hastanesi'nde Allerjik Hastalarda, Allerjen Tesbitinde Spesifik IgE'lerle Deri Testlerinin Karşılaştırılması. XIV. Ulusal Allerji ve Klinik İmmünoloji Kongresi Bildiri Özetleri Kitabı, sayfa: , 4-8 Kasım 2006 (Poster Sunumu)
10. Öztuna D., Elhan A.H.: Madde Yanıt Teorisine Giriş. 9. Ulusal Biyoistatistik Kongresi Bildiri Özetleri Kitabı, sayfa: 24, 5-9 Eylül 2006 (Sözlü Sunum).
11. Köse S.K., Öztuna D., Tüccar E.: Çoklu Testler için Düzeltme Yöntemleri. 9. Ulusal Biyoistatistik Kongresi Bildiri Özetleri Kitabı, sayfa: 53, 5-9 Eylül 2006 (Sözlü Sunum).
12. Kaskatı O.T., Öztuna D., Elhan A.H.: İki Sonuçlu Hale Getirilmeksizin Sürekli Bir Sonuç Değişkeni için Odds Oranının Hesaplanması. 9. Ulusal Biyoistatistik Kongresi Bildiri Özetleri Kitabı, sayfa: 47, 5-9 Eylül 2006 (Sözlü Sunum).
13. Kuyumcu G., Kocaman S., Saltık A., Öztuna A.: Beck Depresyon Ölçeği ile Gebelerde Depresyon Değerlendirmesi. X. Ulusal Halk Sağlığı Kongresi Kongre Kitabı, sayfa: 181, 6-8 Eylül 2006 (Poster Sunumu).

14. Göğüş Ç., Özayar S., Süer E., Öztuna D., Türkölmez K., Baltacı S., Bedük Y.: PT2N0 Böbrek Tümörlü Hastalarda Boyutlara Göre Prognozun Değerlendirilmesi. 19. Ulusal Üroloji Kongresi Kongre Özet Kitabı, sayfa:369, 10-15 Haziran 2006 (Poster Sunumu).
15. Göğüş Ç., Özayar S., Süer E., Öztuna D., Şafak M., Küpeli S.: Lokalize Hücreli Kanserde Histolojik Sınıflandırmanın Prognoz Üzerindeki Etkisi: Şeffaf Hücreli ve Papiller Tip Arasında Fark Var mı? 19. Ulusal Üroloji Kongresi Kongre Özet Kitabı, sayfa:367, 10-15 Haziran 2006 (Poster Sunumu).
16. Erden Ş., Ateş A., Başkal N., Öztuna D.: Obez Bireylerin Obezite Eğitimi ve Hastalıkları İle İlgili Sorunlarının Değerlendirilmesi. 42. Ulusal Diyabet Kongresi Özet Kitabı, sayfa: 67, 10-14 Mayıs 2006 (Poster Sunumu).
17. Özbek U., Öztuna D., Doğanay B., Köse S.K.: Tekrarlı Ölçümlerin Eğri Altında Kalan Alan Yardımıyla Analizi. VIII. Ulusal Biyoistatistik Kongresi Bildiri Kitabı, sayfa: 283-284, 20-22 Eylül 2005 (Sözlü Sunum).
18. Öztuna D., Elhan A.H., Atakurt Y.: Dört Farklı Normallik Testinin Tip I Hata ve Güç Bakımından Değişik Dağılımlarda İncelenmesi. VIII. Ulusal Biyoistatistik Kongresi Bildiri Kitabı, sayfa: 446-447, 20-22 Eylül 2005 (Poster Sunumu).
19. Akmeşe H.E., Gündüz K, Günalp İ., Öztuna D.: Siliyokoroid Melanomlarında Ru-106 Plak Radyoterapisi. 39. Oftalmoloji Kongresi Özet Kitabı, sayfa: 128, 17-21 Eylül 2005 (Sözlü Sunum).
20. Akmeşe H.E., Gündüz K, Günalp İ., Öztuna D.: Koroid Melanomlarında Kombine Ru-106 Plak Radyoterapi ve Transpupiller Termoterapi. 39. Oftalmoloji Kongresi Özet Kitabı, sayfa: 285, 17-21 Eylül 2005 (Poster Sunumu).
21. İleri (Akçaboy) M., Özçakar Z.B., Acar B., Yüksel S., Gökmen D, Ekim M., Yalçınkaya F.: Tedaviye Dirençli Nefrotik Sendromlu Çocuklarda Antihiperlipidemik Tedavi. 41. Türk Pediatri Kongresi Kongre Kitabı, sayfa: 547, 22-25 Haziran 2005 (Sözlü Sunum).
22. Yüksel S., Ekim M., Özçakar Z.B., Acar B., Gökmen D., Yalçınkaya F.: İlk Ateşli İdrar Yolu Enfeksiyonu Sonrası Böbrekteki Hasar Riskine Yaş ve Vezikoüreteral Reflünün Etkisi. 41. Türk Pediatri Kongresi, sayfa: 154, 22-25 Haziran 2005 (Sözlü Sunum).
23. Gökmen H., Öztuna D., Küçükler C., Güllü S., Erdoğan M.F., Emral R., Çorapçioğlu D., Gedik V., Uysal A., Kamel N., Başkal N.: Diyabetiklerin Diyabet Eğitimi ve Hastalıkları ile İlgili Sorunlarının Eğitim Hemşiresi Anketi ile Değerlendirilmesi. 41. Ulusal Diyabet Kongresi Özet Kitabı, sayfa: 120, 11-15 Mayıs 2005 (Sözlü Sunum).
24. Öztuna D., Genç Y.: Kümelenmiş İki Sonuçlu Verilerde Ki-Kare Testinin Uygulanması. 4. İstatistik Kongresi Bildiri ve Poster Özetleri Kitabı, sayfa: 294-295, 8-12 Mayıs 2005 (Sözlü Sunum).
25. Yüksel S, Ekim M., Acar B., Özçakar B., Gökmen D., Yalçınkaya F.: Bir Yaş Altı Çocuklarda İlk Ateşli İdrar Yolu Enfeksiyonu. 21. Ulusal Nefroloji, Hipertansiyon, Diyaliz ve Transplantasyon Kongresi Bildiri Özet Kitabı, sayfa: 66, 2004 (Poster Sunumu).
26. Genç Y., Gökmen D., Tüccar E.: Kümelenmiş (İki Sonuçlu) Verilerde Duyarlılık ve Seçicilik Ölçütlerinin Hesaplanması. 4. İstatistik Günleri Sempozyumu Program ve Özetler Kitabı, sayfa: 101, 20-21 Mayıs 2004 (Poster Sunumu).
27. Gökmen D., Genç Y., Atakurt Y.: Kümelenmiş (İki Sonuçlu) Verilerde McNemar Testinin Uygulanması. 4. İstatistik Günleri Sempozyumu Program ve Özetler Kitabı, sayfa: 113, 20-21 Mayıs 2004 (Poster Sunumu).
28. Gökmen D., Toktamış Ö.: Bant Genişliği Seçiminde Kullanılan Yöntemlerin Simetrik ve Simetrik Olmayan Dağılımlarda Karşılaştırılması. İstatistik Araştırma Sempozyumu Bildiri Özetleri Kitabı, sayfa: 9-10, 2002 (Sözlü Sunum).

## **KENDİ BİLİM ALANINDA YAZILAN TÜRKÇE KİTAP BÖLÜMLERİ**

1. Elhan A.H., Gökmen D., Kuzu M.A.: Araştırma Yöntemleri ve Biyoistatistik. Cerrahi Araştırma, (s. 45-57) Ed.: Cantürk N.Z., Sayek İ., Nobel Tıp Kitabevleri, 2005.

## **BİLİM ALANI İLE İLGİLİ YAYINLANMIŞ KİTAP BÖLÜMÜ ÇEVİRİLERİ**

1. Elhan A., Öztuna D.: Moleküler Filogenetik Analizi. Moleküler Mikrobiyoloji, (s. 161-180) Ed.: Persing D.H., Çev.Ed.: Tekeli A., Ustaçelebi Ş., Palme Yayınları, 2006.

## **VII- Bilimsel Etkinlikleri**

### **PROJELER**

1. Türkiye’de Eğitimin Finansmanı ve Eğitim Harcamaları Bilgi Sistemi Projesi (e-Okul Bütçeleri Projesi). TÜBİTAK 1007 Kamu Kurumları Araştırma ve Geliştirme Programı, Proje Araştırmacı, 2.200.000 YTL, 30 Aralık 2006
2. Kas-İskelet Sorunlarının Özürlülük Değerlendirmede Bilgisayar Uyarlamalı Test Yönteminin Uygulanması. Ankara Üniversitesi Bilimsel Araştırma Projeleri Birimi, Proje Araştırmacı, 14.300 YTL, 2007.
3. Çürük Aktif Çocuklarda Okluzal Çürük Teşhisinde Farklı Yöntemlerin Karşılaştırılması. TÜBİTAK SBAG-HD-79 (105S339) sayılı Hızlı Destek Programı, Proje Araştırmacı, 4.000 YTL, Ocak 2006-Ocak 2007. Proje yürütücüsü: Prof. Dr. Leyla DURUTÜRK
4. Need Analysis for Catch-Up Education Project (Milli Eğitim Bakanlığı ve UNICEF TURKEY, EU FUNDED SC/2005/0460-01, Proje Uzmanı, Şubat 2006-Agustos 2006) 70.284\$
5. The WHO-Care for Development Intervention; A Study to Test Effect of Intervention on Maternal Knowledge and Attitudes Regarding Child Development (Ankara Üniversitesi Çocuk Sağlığı ve Hastalıkları Anabilim Dalı Gelişimsel Pediatri Ünitesi ve WHO, 03011 HNI, İstatistik Uzmanı, 2004)
6. Promoting Child Development: Yale-Ankara Collaboration (Ankara Üniversitesi Çocuk Sağlığı ve Hastalıkları Anabilim Dalı Gelişimsel Pediatri Ünitesi ve NIH, NIH R21 TW006678-01; CFDA Number 93.9899, İstatistik Uzmanı, Mart 2003)

## **EĞİTİMCİ OLARAK YER ALINAN KURS VE ÇALIŞTAYLAR**

1. Göğüs Hastalıkları Hekimleri İçin Temel Biyoistatistik Kursu, Toraks Derneği 8. Yıllık Kongresi, Kemer/ANTALYA, 27 Nisan 2005.
2. Klinik Araştırma Metodolojileri Kursu (Uygulamalı İstatistik), National Institute of Health (NIH) "Çocuk Gelişiminin Desteklenmesi: Yale-Ankara İşbirliği" Projesi Gelişimsel Pediatri Ünitesi Eğitim Programları-II, ANKARA, 22 Ekim 2005.

## **VIII- Diğer Bilgiler**

### **KATILIMCI OLARAK YER ALINAN KURS VE ÇALIŞTAYLAR**

1. Intermediate Rasch Course, Psychometric Laboratory for Health Sciences University of Leeds in the UK, Leeds-İNGİLTERE, 14-17 Mayıs 2007.
2. Klinik Araştırma Metodolojileri Kursu, National Institute of Health (NIH) "Çocuk Gelişiminin Desteklenmesi: Yale-Ankara İşbirliği" Projesi Gelişimsel Pediatri Ünitesi Eğitim Programları-II, ANKARA, 19 Ekim 2005.

## **BİYOİSTATİSTİK DANIŞMANLIĞI BULUNAN DERGİLER**

1. Türkiye Klinikleri Oftalmoloji Dergisi
2. Romatizma Dergisi
3. Journal of Physical Medicine and Rehabilitation Sciences