

**TC.
ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ FAKÜLTESİ
ÖLÇME VE DEĞERLENDİRME ANABİLİM DALI**

**TIP EĞİTİMİNDE GELİŞİM SINAVI SORU BANKASI
OLUŞTURULMASI VE BENZETİM VERİLERİ İLE BİLGİSAYAR
UYARLAMALI TEST UYGULAMASI**

DOKTORA TEZİ

Dr. Ayşen Melek AYTUĞ KOŞAN

Ankara

Haziran 2013

**TC.
ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ FAKÜLTESİ
ÖLÇME VE DEĞERLENDİRME ANABİLİM DALI**

**TIP EĞİTİMİNDE GELİŞİM SINAVI SORU BANKASI
OLUŞTURULMASI VE BENZETİM VERİLERİ İLE BİLGİSAYAR
UYARLAMALI TEST UYGULAMASI**

DOKTORA TEZİ

Dr. Ayşen Melek AYTUĞ KOŞAN

**Danışman
Prof. Dr. Nizamettin KOÇ**

**Eş Danışman
Prof. Dr. Atilla Halil ELHAN**

**Ankara
Haziran 2013**

JÜRİ ÜYELERİNİN İMZA SAYFASI

Eğitim Bilimleri Enstitüsü Müdürlüğüne,

Bu çalışma jürimiz tarafından Ölçme ve Değerlendirme Anabilim Dalında
DOKTORA TEZİ olarak kabul edilmiştir.

Başkan  (Danışman)

Prof. Dr. Nizamettin KOÇ

 (Eş Danışman)

Prof. Dr. Atilla H. ELHAN

Üye 

Prof. Dr. Ezel TAVŞANCIL

Üye 

Prof. Dr. Selahattin GELBAL

Üye 

Doç. Dr. Derya ÖZTUNA

Üye 

Yrd. Doç. Dr. Deniz GÜLLEROĞLU

ONAY

Yukarıdaki imzaların, adı geçen öğretim üyelerine ait olduğunu
onaylarım.

.../.../2013

Prof. Dr. İsmail GÜVEN

Enstitü Müdürü

ÖNSÖZ

Gerek doktora eğitimim boyunca, gerekse bu çalışmanın yürütülmesinde desteğini esirgemeyen, çalışmamı bilgi, görüş ve önerileriyle yönlendiren ve çalışmam boyunca hep yanımda olduğunu hissettiğim değerli danışmanım Prof. Nizamettin KOÇ'a

Çalışmam süresince değerli bilgi, görüş ve deneyimlerini benimle paylaşan, bana rehberlik eden, beni motive eden ve bilgisinin yanı sıra zamanını cömertçe sunmaktan geri kalmayan değerli eş danışmanım Prof. Dr. Atilla Halil ELHAN'a

Bu çalışmada ve desteğine ihtiyaç duyduğum her durumda bilgisini, çabasını ve zamanını esirgemeyen, beni destekleyen, çalışma dışında da yakınımda bulunmasından keyif aldığım, değerli arkadaşım Doç. Dr. Derya ÖZTUNA'ya

Bilgisi ile olduğu kadar, üretmeye ve gelişmeye olan inancı ile de hep desteğini hissettiğim, dostluğu ve aydınlık çehresiyle hep çok yakınımda olan ve gelecekte de desteğini daima bulacağımdan emin olduğum değerli arkadaşım Uz. Dr. Meral DEMİRÖREN'e

Hayatımın her alanında olduğu gibi, doktora eğitimim ve bu uzun çalışma süresi boyunca bana sabır ve sevgi ile destek olan, tüm söylemek istediklerimi kelimelerle ifade edemeyeceğime düşündüğüm sevgili eşim Doç. Dr. Murat KOŞAN'a

Sonsuz teşekkürlerimi sunarım.

Ayrıca araştırma süresince yardımlarını gördüğüm isimlerin sayamadığım herkese teşekkürlerimi sunmayı borç bilirim.

ÖZET

TIP EĞİTİMİNDE GELİŞİM SINAVI SORU BANKASI OLUŞTURULMASI VE BENZETİM VERİLERİ İLE BİLGİSAYAR UYARLAMALI TEST UYGULAMASI

Doktora, Ölçme ve Değerlendirme Anabilim Dalı

Tez Danışmanı: Prof. Nizamettin KOÇ,

Eş Danışman: Prof Dr. Atilla H. ELHAN

Mayıs 2013, IX+117 Sayfa

Bu araştırmada, Ankara Üniversitesi Tıp Fakültesinde yürütülmekte olan gelişim sınavı için bir soru bankası oluşturulması ve bu soru bankasını kullanarak BUT yönteminin uygulanabilirliğinin ortaya konulması amaçlanmıştır.

Soru bankasının oluşturulmasında iki kategorili yanıtlar için Rasch modeli kullanılmış, modelin varsayımlarının karşılanıp karşılanmadığını belirlemeye yönelik çalışmalar yapılmış ve 103 maddeden oluşan tek bir boyut içeren soru bankası elde edilmiştir. Soru bankasında bulunan 103 maddenin/soru bankasının Birey Ayırsama İndeksi ve KR-20 güvenirlik katsayısı değerleri 0.77 olarak hesaplanmıştır. Soru bankasında yer alan maddelerin ve öğrencilerin θ düzeyi boyunca dağılımı incelendiğinde θ 'nın her düzeyi için yeterli sayıda madde olduğu sonucuna ulaşılmıştır.

Çalışma bulguları, Gelişim Sınavı kağıt kalem testi uygulamaları ve BUT uygulamasının ile elde edilen yetenek kestirimleri arasındaki korelasyonun (ortalaması 0 ve varyansı 1 ve ortalaması 0 varyansı 3 normal dağılım; 0.3, 0.4, 0.5, 0.548 standart değerler için) yüksek olduğunu göstermektedir.

Ortalaması 0 ve varyansı 1 olan dağılımdan 1000 birey için türetilen yanıtlar ile gerçekleştirilen BUT'da kullanılan madde sayısı ortalaması

standart hatanın farklı düzeyleri için 11 ile 45 arasında değişmektedir. BUT uygulamalarında ortalama 14 madde ile kestirim yapılabilmektedir. BUT uygulamalarında, kağıt kalem uygulamalarına göre madde sayısında %56.3-86.4 arasında azalma sağlandığı gösterilmiştir.

Ortalaması 0 ve varyansı 1 olan dağılımdan 1000 birey için türetilen yanıtlar ile gerçekleştirilen BUT'da kullanılan madde sayısı ortalaması standart hatanın farklı düzeyleri için 12 ile 75 arasında değişmektedir. BUT uygulamalarında, kağıt kalem uygulamalarına göre madde sayısında %27.6-88.3 arasında azalma sağlandığı bulunmuştur.

Araştırmanın sonuçları, Rasch yöntemi ile Gelişim Sınavı için uygun bir soru bankası hazırlanabileceği, bu soru bankası kullanılarak Gelişim Sınavının, BUT yöntemi ile uygulanabileceği ve güvenilir sonuçlar elde edilebileceğini göstermektedir. BUT uygulaması ile sınava giren kişilerin sınavın klasik formunda bulunandan çok daha az sayıda madde ile yetenek kestirimleri yapılabileceği gösterilmiştir.

SUMMARY

DEVELOPING AN ITEM BANK FOR PROGRESS TEST AND APPLICATION OF A COMPUTERIZED ADAPTIVE TESTING BY SIMULATION IN MEDICAL EDUCATION

Department of Measurement and Evaluation

Advisors: Prof. Nizamettin KOÇ, Prof Dr. Atilla H. ELHAN

May 2013, IX+117 Pages

In this research, it has been aimed to develop an item bank for the progress test carried out at the Faculty of Medicine at Ankara University and to present the applicability of CAT method using the item bank.

While developing the item bank, for dichotomus items Rasch model was used, the studies for meeting the assumptions were carried out and after all a unidimensional item bank involving 103 items was obtained. Person Separation Index-(PSI) and KR-20 of the item bank was calculated as 0.77. Considering the items in the bank and the distribution of students through θ level, it was concluded that there were adequate number of items for each level of θ .

The findings of research have indicated that the correlation (average mean is 0, variance 1 and average mean 0 variance 3 normal distribution; for standard values 0.3, 0.4, 0.5, 0.548) between the θ estimations obtained from the paper-and-pencil progress test and CAT applications is high.

The average mean of items used in CAT, which was carried out with derived responses for 1000 individuals from the distribution of mean with 0 and variance with 1, varies between 11 and 45 for various levels of standard error. In BUT applications, it is possible to estimate with 14 items. It has been observed that there is a decrease between 56.3% and 86.4% in CAT applications for the number of items compared with paper-and-pencil tests.

The average mean of items used in CAT, which was carried out with derived responses for 1000 individuals from the distribution of mean with 0 and variance with 3, varies between 12 and 75 for various levels of standard error. It has been observed that there is a decrease between 27.6% and 88.3 %in BUT applivations for the number of items compared with paper-and-pencil tests.

The results of research have indicated that it is possible to develop a proper item bank for the Progress Test using Rasch Model and also using this item bank, the progress test can be applied with CAT method and it is likely to get reliable results. Besides, it has been found out that the theta estimations of individuals taking the progress test can be done with quite less items in CAT application when compared with the classical form of the test.

İÇİNDEKİLER

JÜRİ ÜYELERİNİN İMZA SAYFASI	ii
ÖNSÖZ	iii
ÖZET	iv
SUMMARY	vi
ÇİZELGELER LİSTESİ	x
ŞEKİLLER LİSTESİ	xi
BÖLÜM I	1
GİRİŞ	1
Problem	1
Araştırmanın Amacı	19
Önem	19
Varsayımlar	20
Sınırlılıklar	20
Tanımlar	20
Kısaltmalar	22
BÖLÜM II	23
KURAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR	23
Gelişim Sınavı	23
Soru Bankası	31
Bilgisayar Uyarlamalı Testler	36
Tıp Eğitiminde Bilgisayar Uyarlamalı Test	48
BÖLÜM III	52
YÖNTEM	52
Araştırma Modeli	52
Çalışma Grubu	52
Veri Toplama Aracı ve Verilerin Toplanması	52
BÖLÜM IV	73
BULGULAR VE YORUMLAR	73
1.1. Soru Bankasının psikometrik özellikleri (Soru Bankasının İçsel Yapı Geçerliliği - Internal Construct Validity ve Güvenirlik)	73

2.1. Simüle uygulama için veri türetilmesi ve simüle verilerle BUT uygulama (Soru Bankalarının Dış Geçerliliği- External Validity)	82
BÖLÜM V	92
SONUÇ VE ÖNERİLER.....	92
Sonuçlar	92
İleri Araştırma Önerileri.....	95
KAYNAKÇA	97
EKLER	107

ÇİZELGELER LİSTESİ

Çizelge 1	Ankara Üniversitesi Tıp Fakültesi Gelişim Sınavına Katılan Öğrenci Sayısının Yıllara Göre Dağılımı ve Toplam Öğrenci Sayısı İçindeki Yüzdesi	53
Çizelge 2	2011 Gelişim Sınavına Katılan Öğrenci Sayısının Dönemlere Göre Dağılımı	74
Çizelge 3	2011-2012 Eğitim Öğretim Yılı Gelişim Sınavı “Genel Tıp” Soru Bankasının Rasch Modeline Uyumu	75
Çizelge 4	BUT Uygulaması ile Öğrencilerin θ Düzeyi Kestirimleri için Belirlenmiş Standart Hata ve Güvenirlilik Düzeyleri için Tanımlayıcı İstatistikler (Ortalaması 0, Varyansı 1 Olan Normal Dağılım İçin)	83
Çizelge 5	BUT Uygulaması ile Öğrencilerin θ Düzeyi Kestirimleri İçin Belirlenmiş Standart Hata ve Güvenirlilik Düzeyleri için Tanımlayıcı İstatistikler (Ortalaması 0, Varyansı 3 Olan Normal Dağılım İçin)	86
Çizelge 6	Gelişim Sınavı Uygulamaları Kullanılma Durumu	107
Çizelge 7	MFF Gösteren Maddelerin Dağılımı	112
Çizelge 8	103 Maddelik Soru Bankasında Yer Alan Maddelerin Sistem ve Bilim Alanlarına Göre Dağılımı	115

ŞEKİLLER LİSTESİ

Şekil 1	a) İki Parametrel Modelde Üç Maddenin Karakteristik Eğrileri, b) Üç Parametrel Modelde Üç Maddenin Karakteristik Eğrileri	13
Şekil 2	Rasch Modelinde Üç Maddenin Madde Karakteristik Eğrileri	14
Şekil 3:	Soru Bankası Geliştirme Basamakları	33
Şekil 4	Soru Bankasındaki Maddelerin Kalibrasyonu için Oluşturulmuş Alt Test/Alt Grupları İçeren Modeller.	35
Şekil 5:	Bilgisayar Uyarlamalı Test Değerlendirme Modeli	38
Şekil 6:	Bilgisayar Uyarlamalı Test Süreci	43
Şekil 7	Madde I0001 Madde Analizi	60
Şekil 8	Madde I0020 Madde Analizi	61
Şekil 9	Bland-Altman Grafiği Örneği	72
Şekil 10	“Genel Tıp” Final Soru Bankasındaki Maddelerin ve Öğrencilerin θ Düzeyi Boyunca Dağılımı	78
Şekil 11	Ortalaması 0, Varyansı 1 Olan Normal Dağılım için Bland- Altman Grafikleri; (a) SH=0.30, (b) SH=0.40, (c) SH=0.50, (d) SH=0.548	85
Şekil 12	Ortalaması 0, Varyansı 1 Olan Normal Dağılım için Bland- Altman Grafikleri; (a) SH=0.30, (b) SH=0.40, (c) SH=0.50, (d) SH=0.548	89

BÖLÜM I

GİRİŞ

Bu bölümde, araştırmaya ilişkin problem durumu, araştırmamanın amacı, önemi, varsayımları, sınırlılıkları ve tanımlar yer almaktadır.

Problem

Öğrenme kavramı çağımızın değişen ve kendini yenileyen dinamiği içinde gelişmektedir. Bu bağlamda, tıp eğitimi de toplumun sağlığını geliştirecek temel yeterliklere sahip hekimleri yetiştirme yönünde değişmekte ve gelişmektedir.

Tıp eğitimi ile konuya özel içerik hedeflerinin kazandırılmasının yanı sıra, öğrencilere kendi kendilerine öğrenebilme, bilgiyi bağımsız biçimde elde edip analiz edebilme, problem çözme, eleştirel düşünme ve akıl yürütme temel becerilerinin de kazandırılması hedeflenmektedir. Bilgi çağının da hedeflediği bu temel becerilerin, geleneksel eğitim ortamlarında gerçekleşmesinin güç olması nedeniyle yeni öğretim yaklaşımları ve beraberinde yeni değerlendirme yaklaşımları gündeme gelmektedir.

Öğrenci başarısının değerlendirilmesi, son yıllarda önemli değişim gösteren süreçlerden biri olmuştur. Bu bağlamda, “test etme kültürü”nden “değerlendirme kültürü”ne bir geçiş yaşanmaktadır. Eğitim ile ölçme ve değerlendirme arasındaki ilişki “ürün değerlendirme”den “süreç değerlendirme”ye doğru kaymıştır. Buna paralel olarak değerlendirmenin işlevi ve biçimi de değişmiş ve zenginleşmiştir. Gerçek hayat ile uyumlu ve çok çeşitli değerlendirme araçları kullanılması giderek yaygınlaşmaktadır (Birenbaum ve Dochy, 1996). Eğitici merkezlikten öğrenci merkezli eğitime ve değerlendirmeye doğru yönelim hız kazanmış, öğrencilerin ihtiyaç ve beklentilerine verilen önem de artmıştır.

Bu süreçte “öğrenmenin değerlendirilmesi” kadar “öğrenme için değerlendirme”nin önemine yapılan vurgu da artmıştır. Ölçme değerlendirme sistemini oluştururken farklı amaçları içeren ölçme değerlendirme süreçlerinin sistem içinde yer alması önemli görülmektedir (Hammerman, 2009). Bir eğitim ve öğretim sürecinin belirli aşamalarında belirli amaçlarla yapılan değerlendirmeler aşağıda özetlenmiştir (Tekin, 2009).

1. *Programa girişte yapılan değerlendirme (Tanıma-yerleştirmeye yönelik değerlendirme):* Öğrencinin programa başlamadan önce, ön öğrenmelerini (bilişsel, duyuşsal ve psikomotor) tanımlamak amacıyla yapılan değerlendirmedir. Öğrencilerin giriş düzeylerinin bilinmesi, gelişimlerinin değerlendirilmesine temel oluşturur.

2. *Program sürecinde yapılan değerlendirme (Biçimlendirme-yetiştirmeye yönelik (formatif) değerlendirme):* Öğrencilerin öğrenmedeki güçlüklerini ortaya çıkarmak ve gerekli düzeltmeleri yapmak için eğitim süresince yapılan değerlendirmedir. Programın zayıf ve güçlü yönlerini ortaya çıkararak sürekli geribildirim sağlar ve iyileştirici önlemlerin alınması için bir kontrol sistemi oluşturur.

3. *Programın sonunda yapılan değerlendirme (Değer biçmeye-belgelemeye yönelik (summatif) değerlendirme):* Program sonunda öğrencilerin kazandığı davranış, özellik ve becerileri ölçmek amacıyla yapılan değerlendirmedir. Bu değerlendirme ile öğrencilere istenen davranışları kazandırma açısından eğitim programının yeterli olup olmadığı hakkında yargıya varılabilir. Başarı veya yeterlik testleri ile öğrencilere kazandırılması hedeflenen tüm özellikler test edilmeye çalışılır.

Bound (2007) değerlendirmede öğrenmeyi birincil amaç olarak merkeze koymaktadır. Bu yaklaşımla öğrenciler kendi öğrenmelerine ilişkin yargıda bulunabilmeli ve bu yargının sonuçlarından daha sonraki öğrenmeleri için faydalanabilmelidirler. Bu amaca hizmet etmek üzere biçimlendirmeye yönelik değerlendirmeler yapılmaktadır.

“Formatif” ve “summatif” kavramları ilk olarak 1967’de Michael Scriven tarafından değerlendirme için veri toplamanın amaçları ve bu toplanan verileri kullanılırken izlenen yollar arasındaki farklılığa vurgu yapmak üzere kullanılmıştır (Greenstein, 2010). Bununla birlikte, biçimlendirmeye yönelik değerlendirmeyi eğitimde ölçme değerlendirmede ilk uygulayan kişi Benjamin Bloom olmuştur. Bloom, öğrencilerin belirli bir düzeydeki öğrenim hedeflerinde yeterlilik göstermeden bir üst düzeydeki yeterlikleri öğrenmeye geçmemeleri temeline dayanan tam öğrenme (mastery learning) kavramını oluştururken, biçimlendirmeye yönelik değerlendirmeyi merkeze almıştır. Bloom biçimlendirmeye yönelik değerlendirmenin temel iki bileşeni olduğunu belirtmektedir: Bunlardan birincisi öğrencilere geribildirim sağlama, ikincisi de öğrenmenin önemli bileşenleri için düzeltme/geliştirme fırsatları yaratmak.

Bu kavramların ortaya konmasından bu yana biçimlendirmeye yönelik değerlendirmeye ilişkin pek çok çalışma yapılmış, 1998 yılında Paul Black ve Dylan William’ın 250 araştırma ile yapmış oldukları meta analiz biçimlendirmeye yönelik değerlendirmenin yaygınlaşmasına önemli katkı sağlamıştır (Greenstein, 2010).

Biçimlendirmeye yönelik değerlendirme için alan yazında pek çok tanım bulmak mümkündür. Biçimlendirmeye yönelik değerlendirme “öğrenme hakkında sürekli kanıt toplamaya yönelik sistematik bir süreç, eğitim süresince eğiticiye ve öğrenciye geribildirim sağlamak için değerlendirmenin tanı koymak üzere kullanılması” olarak tanımlanmaktadır (Heritage, 2007, Boston, 2002).

Biçimlendirmeye yönelik değerlendirme öğrenme sürecini izleme, eğiticiye ve öğrenciye öğrenme hedeflerine ulaşmadaki gelişimlerine ilişkin geribildirim sağlama, öğrenmede süreklilik ve ek öğrenmeler için motivasyon sağlama ve bu yolla öğrenmeyi süreç içinde desteklemek amaçlarına hizmet eder.

Biçimlendirmeye yönelik değerlendirme standartlar, bağlam/içerik ve değerlendirme arasında bağlantı kurulmasını sağlar; değerlendirmeyi eğitimin bir parçası haline getirir ve eğitimsel kararlar almaya rehberlik eder. Tıp

eğitiminde ölçme değerlendirmede sisteminde yer alması gereken biçimlendirmeye yönelik değerlendirme öğrenci merkezli eğitimin önemli bir parçasıdır. Bu amaçla yapılmış olan değerlendirmelerden elde edilen puanlar, bireylerin başardı-başaramadı kararlarını oluşturmaktan çok, onların öğrenmelerinde ortaya çıkan aksaklıkların giderilmesi amacıyla kullanılır (Heritage, 2007, Boston, 2002).

Tıp eğitiminde biçimlendirmeye yönelik değerlendirme ölçme değerlendirme süreci içinde temel bir bileşeni oluşturmalıdır. Bu bağlamda, önemli bir biçimlendirmeye yönelik değerlendirme yöntemi olarak Gelişim Sınavı (Progress Test), gün geçtikçe daha önemli bir yer edinmektedir.

Gelişim Sınavı

Tıp Eğitiminde Gelişim Sınavı bir eğitim programında farklı düzeylerde bulunan bir grup öğrenciye aynı maddelerden oluşan bir testin uygulandığı bir sınavdır. Gelişim sınavı, herhangi bir blok, kurul, staj ya da dönem sonuna özel olarak yapılmayan, öğrencilerin bulundukları seviyeyi ve gelişimi ölçmeye yönelik, fakülte'deki tüm öğrencilerin aynı anda, aynı maddelerle karşılaştıkları bir sınavdır (Freeman, van der Vleuten, Nouns, Ricketts ve Welbourn, 2010).

Gelişim Sınavı, eğitimin sonunda öğrencinin elde etmesi beklenen tüm çekirdek eğitim hedeflerini içerecek şekilde planlanır. Her Gelişim Sınavında maddelerin, tıp eğitiminin tüm evrelerini kapsamaları hedeflenir. Diğer bir ifadeyle maddeler farklı disiplinlere ya da eğitim sürecindeki ünite, blok veya modüllerden herhangi birine göre değil, mezuniyet aşamasında bilinmesi gerekenlere göre düzenlenir. Bir tıp fakültesindeki bütün sınıflar/dönemler, yıl boyunca yapılan gelişim sınavlarına düzenli aralıklarla aynı anda girerler (Rademakers, Cate ve Bar, 2005).

Gelişim Sınavı, bilişsel alan öğrenme hedefleri ile ilişkili bir belirtke tablosuna bağlı kalınarak belirlenen ve eğitim programındaki ilgili disiplinler ile konu alanlarının temsil edildiği bir madde örnekleminde oluşmaktadır.

Her Gelişim Sınavı için aynı belirtke tablosu kullanılmakla beraber, her Gelişim Sınavı farklı maddelerden oluşmaktadır. Gelişim Sınavı belirli aralıklarla düzenli olarak tekrarlanır. Bu yolla öğrencilerin kişisel ilerlemeleri belirlenir (Dent ve Harden, 2005).

Gelişim Sınavında maddeler “fonksiyonel bilgi”yi sorgulamak üzere hazırlanmaktadır. Fonksiyonel bilgi “tıp eğitimi süresince verilen ve mezuniyet sonrası hekimlik uygulamalarında kullanabilecek bilgi veya hekimlik uygulamaları ile doğrudan ilişkilendirilebilir bilgi” olarak tanımlanmaktadır (van der Vleuten, Verwijnen ve Wijnen, 1996 a).

Gelişim Sınavında öğrencilerin, seviyelerine uygun olarak kazanmaları beklenen bilişsel öğrenme hedeflerine ulaşma durumları değerlendirilir. Diğer bir ifadeyle, öğrenciler maddelere, sınava girdikleri tarihe kadar aldıkları eğitim doğrultusunda yanıtlar verebileceklerdir. Sınavda doğru yanıtlanan madde sayısının her geçen yıl giderek artması beklenir (Dent ve Harden, 2005). Böylece, eğitim programının sonunda ulaşılmak istenen hedeflerle ilişkili olarak öğrencilerin kazanımları ile ilgili, geniş kapsamlı, kesitsel (farklı yetenek düzeyindeki öğrencilerin performanslarının karşılaştırıldığı) ve boylamsal (bireylerin belirli aralıklarla performanslarının karşılaştırıldığı) bir profil elde edilir (Freeman ve diğerleri, 2010).

Gelişim Sınavı’nın değerlendirilmesiyle birlikte öğrenciler, performanslarına ilişkin, hem kapsamın bütünü hem de farklı disiplinler ve alt disiplinler düzeyinde bireyselleştirilmiş geribildirim alırlar. Ayrıca bu geribildirimlerle öğrencilere kendi bulundukları dönem ve tüm dönemler bağlamında bulundukları öğrenme düzeylerine ilişkin bilgi sağlanır. Öğrenciler güçlü ve zayıf yönlerine, hangi konulara ağırlık vermeleri gerektiğine ilişkin bilgi sahibi olurlar (Blake ve diğerleri, 1996).

Gelişim Sınavları, Tıp Eğitiminde 1970’li yıllarda Missouri Kansas City Tıp Okulunda (ABD) (Dönem Profili Sınavı-the Quarterly Profile Examination-QPE adı ile) ve Limburg-Maastricht (Hollanda) üniversitesinde aynı zamanda geliştirilmiştir, ancak farklı amaçlarla uygulanmaya başlamıştır (McHarg, Bradley, Chamberlain, Ricketts ve McLachlan, 2005). Daha sonra

benzer test uygulaması Mc Master Üniversitesi (Kanada) Kişisel Profil İndeksi (Personel Profile Index-PPI) adı altında yapılmaya başlamıştır. Gelişim Sınavları halihazırda hemen hemen bütün kıta/bölgelerde uygulanmaktadır (Rademakers ve diğerleri, 2005; Blake ve diğerleri, 1996; Freeman ve diğerleri, 2010).

Hollanda'da gelişim sınavının ortaya çıkmasındaki temel etken Probleme Dayalı Öğrenim (PDÖ) olmuştur. PDÖ'de öğrencilerin kendi öğrenmelerini kendilerinin yönlendirmeleri hedeflenir. Bu hedeflere ulaşılmasında ölçmenin de rolü vardır. Derslerin/modüllerin sonunda yapılan sınavlar öğrencilerin öğrenmelerini, kendi koydukları hedeflerden daha çok yönlendirmektedir. Belirli bir dersin/modülün kapsamına bağlı olmadan ancak eğitimin sonunda ulaşılması planlanan hedefler doğrultusunda uygulanan kapsamlı bir test, öğrencilerin dersin/modülün sonunda yapılan sınava hazırlanmaları dışında da öğrenmelerini, PDÖ oturumlarında ortaya koydukları hedeflerine yönelik çalışmalarını ve derin öğrenme stratejisi geliştirmelerini desteklemektedir (Freeman ve diğerleri, 2010).

Gelişim Sınavı, öğrenci merkezli eğitimin daha sınırlı olduğu geleneksel eğitim programlarında da öğrenci performansını değerlendirmede kullanılmaktadır. Geleneksel eğitim programı uygulayan Missouri Kansas Üniversitesi'nde Gelişim Sınavı öğrenci performanslarını değerlendirme ve karşılaştırma için kullanılmaktadır.

Gelişim Sınavı;

- Bir öğrenci grubunun (kohortunun) ortalama bilgi düzeyini
- Bir öğrenci grubunun (kohortunun) zaman içinde bilgi düzeyindeki değişimini ve o zamana kadar gösterdikleri gelişimini
- Her bir öğrencinin bilgi düzeyini
- Her bir öğrencinin zaman içinde bilgi düzeyindeki değişimi ve o zamana kadar gösterdiği gelişimini
- Her bir test maddesine öğrenci kohortunun verdiği yanıtlarını
- Her bir test maddesine her bir öğrencinin verdiği yanıtı değerlendirmek amacına hizmet eder (Ricketts, Freeman, Pagliuca, Coombes ve Archer, 2010).

- Birden fazla okulda uygulandığı durumda, öğrenci kohortları ve madde bazında okullar arası karşılaştırma yapma olanağı da sağlar (Albano ve diğerleri, 1996).

Özetle, Gelişim Sınavı, öğrenci kohortları, bireysel olarak öğrenciler, tek tek maddeler hatta bazı durumlarda kurumlar bazında karşılaştırmalar yapmaya olanak sağlar.

Gelişim Sınavı dünyanın çeşitli tıp fakültelerinde farklı biçimlerde uygulanmaktadır. Uygulamalar, madde tipleri (çoktan seçmeli, doğru yanlış, açık uçlu maddeler) veya sonuçların kullanılmasına (biçimlendirici, düzey belirleyici, biçimlendirici ve düzey belirleyici bir arada) göre de farklılıklar göstermektedir.

Gelişim Sınavında yer alan maddeler geniş bir davranış örneklemini içerdiğinden, genellikle tek bir doğru cevabın olduğu çoktan seçmeli soru tipinden oluşmaktadır. Bununla beraber, Ulusal Tıp Sınavları Kurulu (National Board of Medical Examiners- NBME-ABD) ile işbirliği yapan İngiltere'deki dört okul ve Amerika'daki iki okul dışında, Gelişim Sınavlarının hemen hemen hepsi kâğıt kalem testi olarak uygulanmaktadır (Freeman ve diğerleri, 2010).

Gelişim Sınavlarının Avantajları

Uzun yılların getirdiği deneyimler ışığında, gelişim sınavının avantajları şu şekilde sıralanabilir (van der Vleuten ve diğerleri, 1996a; Verhoeven ve diğerleri, 2005; De Champlain ve diğerleri, 2010):

- Gelişim Sınavı önemli bir bilgi kaynağı oluşturmaktadır: kesitsel ve boylamsal niteliği, çok kapsamlı olması hem öğrencilere, hem de eğitim programını değerlendiren eğitici ve yöneticiler ile iç ve dış paydaşlara çok miktarda bilgi sağlamaktadır. Gelişim Sınavından elde edilen puanlar, öğrencilerin bilişsel kazanım profillerini belirlemek üzere farklı alt puanlar olarak bir araya getirilebilir (organ- sistem, disiplinler, klinik ve temel bilimler gibi üst disiplin grupları vb). Gelişim Sınavı ayrıca, öğrencilere neler bilmedikleri konusunda da geribildirim sağlar.

- Gelişim Sınavı uzun süreli bilgi birikimini teşvik eder; Gelişim Sınavı kapsamında oldukça farklı konu, sistem/disiplin bulunduğundan, son anda yoğun çalışmayla başarılı olunamaz. Öğrencilerin sürekli ve düzenli çalışmalarını gerektirir. Ayrıca önceki bilgiler periyodik olarak sınıandığından “geç ve unut” yaklaşımını önleyebilir.

- Gelişim Sınavı bütünleme, tekrar sınama (re-sit examination) gereksinimini ortadan kaldırır; her gelişim sınavı, bir anlamda bir önceki gelişim sınavının bütünlemesi sayılabilir.

- Gelişim Sınavı yüksek başarılı öğrencilerin erkenden tespit edilmesini sağlar; bazı öğrenciler eğitim dönemlerinde kendilerinden beklenenden çok daha yüksek performans gösterebilirler. Bu öğrencilere, bireysel performanslarına bağlı olarak eğitim programı kapsamında daha hızlı ilerlemelerini sağlayacak fırsatlar sunulabilir.

- Gelişim Sınavları değerlendirme yöntemlerine ve süreçlerine tutarlılık getirebilir; Gelişim Sınavı, temel olarak her bir ders, modül veya ünitenin öğrenme hedeflerini değil, tüm eğitim programının nihai hedeflerini yansıtır. Eğitim sonundaki nihai kazanımlar değişmediği sürece, eğitim programında yapılan değişiklikler ve kapsam değişiklikleri gelişim sınavı için önem taşımaz.

- Gelişim Sınavları kıyaslama olanağı sağlar; gelişim sınavlarının kapsamı tek bir okul ya da sadece PDÖ müfredatları ile sınırlı olmadığı için farklı üniversite mezunlarının ve farklı eğitim programlarının karşılaştırmalarında kullanılabilir

Araştırmalar Gelişim Sınavının derin öğrenme stratejilerini desteklediğini (Berkel, Nuy ve Geerling, 1995), gelişim örüntülerini ortaya koyabildiğini, (Blake ve diğerleri, 1996; Verhoeven, Verwijnen, Scherpbier ve van der Vleuten, 2002; van Diest ve diğerleri, 2004), özellikle boylamsal bilgiler kullanıldığında ulusal ve uluslararası karşılaştırmalar için uygun olabileceğini göstermektedir (van der Vleuten ve diğerleri, 2004; Muijtjens, Schuwirth, Cohen-Schotanus, Thoben, van der Vleuten, 2008).

Gelişim Sınavının Sınırlılıkları;

- Gelişim Sınavı, eğitimlerinin başında bulunan ilk yıl öğrencileri için zorlayıcıdır. Sınavın, sadece sınırlı bir kısmı (maddelerin yaklaşık % 20'si) eğitim programının ilk yıllarındaki öğrenmelerle ilişkilidir. Bu nedenle Gelişim Sınavı yetenek ölçeğinde alt uçta bulunan ilk yıl öğrencileri arasındaki farklılıkların ortaya konmasında daha az duyarlı olabilir. Ayrıca bu öğrencilerin puanları için hesaplanan yordama geçerliği de daha düşük olabilir.
- Yüksek başarı odaklı olan tıp öğrencileri için maddelerin büyük kısmına cevap verememek moral bozucu ve bıktırıcı olabilir (McHarg ve diğerleri, 2005).
- Uzun testler, yorgunluğa neden olmakta ve öğrencilerin dikkat ve motivasyonlarını azaltabilmektedir.
- Eğitimin sonundaki nihai hedefler doğrultusunda kapsamın tanımlanması, ilişkili konu ve alanların belirlenmesi disiplinler ve konu alanlarının temsil edildiği bir madde örnekleminin oluşturulması zordur. Alt alanlar için elde edilen puanların güvenilirliğinde, her alt alana/ilişkili konu için sorulabilecek madde sayısının kısıtlılığı nedeniyle sıkıntı yaşanabilir.
- Gelişim Sınavının hazırlanması, yürütülmesi ve değerlendirilmesi merkezi bir düzenleme gerektirmektedir.
- Soru bankasının oluşturulması, geliştirilmesi, sürdürülmesi, korunması zordur ve yoğun çaba gerektirir.
- Benzer madde güçlüklerine sahip sınavlar hazırlamanın güç olması da diğer önemli sınırlılıklarındandır (van der Vleuten ve diğerleri, 1996a, Muijtjens ve Wijnen, 2010).

Gelişim Sınavı, soru bankası oluşturma, test geliştirme aşamalarında ve sınavın uygulanması sırasında gerekli yer, zaman, insan gücü ve araç gereç ele alınarak, en üst düzeyde öğrenci katılımını sağlayacak şekilde

planlanmalıdır. Bunlar göz önüne alındığında, sınavın tasarlanması, hazırlanması, sınav güvenliğinin sağlanması, sınavın yürütülmesi ve sonuçlarının değerlendirilmesinin önemli ölçüde, insan gücü, zaman, para v.b. kaynaklar gerektireceği açıktır (Ricketts ve diğerleri, 2010). Bunlar da her üniversite için değişen derecelerde sınırlılıklar olarak ortaya çıkabilir.

Gelişim Sınavı, ilk olarak 1970'li yıllarda uygulanmaya başlamakla beraber Tıp Fakültelerinde yaygın olarak kullanılmaya başlaması uzun zaman almıştır. Gelişim Sınavının sunduğu olanakların çok iyi anlaşılammış olması bunun nedenlerinden biri olabilir. Bir diğer önemli neden de yukarıda daha önce bahsedildiği gibi bu sınavın hazırlanması ve uygulanmasının lojistik yönden zor ve maliyetli olması ve testin geliştirilmesi, uygulanması ve değerlendirilmesinin oldukça fazla çaba ve zaman gerektirmesidir. Bu sınırlılıklara birlikte son yıllarda Gelişim Sınavını uygulayan tıp fakültesi ve kurumların sayısı artmakta, böylece gelişim sınavı tıp eğitiminde değerlendirmenin önemli bir parçası haline gelmektedir (Freeman ve diğerleri, 2010).

Gelişim Sınavı, bireylerin zaman içinde gelişimlerini izlemede önemli bir araçtır. Ancak, tekrarlı testler ile elde edilen bireylere ait tekrarlı puanların, aynı ölçek üzerinde değerlendirilmesine ilişkin yeterli çalışmanın olmamasının, sınavın önemli sınırlılıkları arasında yer aldığı belirtilmektedir (Langer ve Swanson, 2010).

Diğer yandan, gelişim sınavları için, uygun maddelerden oluşan güvenilir, geçerli sınavlar hazırlamaya ve paralel testler oluşturmaya olanak verecek şekilde, iyi hazırlanmış maddelerden oluşan bir soru bankasının olmasının gerekliliği de önemle vurgulanmaktadır (van der Vleuten ve diğerleri, 1996a; Downing ve Haladayne, 2004; Ricketts ve diğerleri, 2010).

Soru bankası

ilk olarak 1960'lı yıllarda İngiltere'de eğitimciler, ortak bir referans ölçek üzerinde öğrencilerin başarılarını değerlendirmede kullanabilecekleri, MTK ile kalibre edilmiş sorulardan oluşan geniş bir soru bankası sunmak üzere

tasarlanmıştır. Soru bankası eğitimde, ölçme çalışmalarında yaygın olarak, kağıt-kalem testi ve Bilgisayar Uyarlamalı Test (BUT) uygulamalarında kullanılmaktadır (Thissen, Reeve, Bjorner ve Chang, 2007).

BUT uygulamaları için soru bankası, testin her aşamasında bireyler için en uygun sorunun seçilmesine olanak sağlayacak şekilde ve ölçülmek istenen yeterliği sınavabilen, kalibrasyonu yapılmış, psikometrik özellikleri tanımlanmış yeterli sayıda sorudan oluşmalıdır (Molina, 2008).

İyi bir soru bankası, geçerli (ölçülen yapının tüm özelliklerini içeren) ölçme sürecinde yüksek ölçme doğruluğu sağlayacak yeterli sayıda soruya sahip olmalıdır. Soru bankasındaki sorular iyi bir soruda olması gereken niteliklere sahip olmalı ve farklı alt gruplar için aynı şekilde fonksiyon göstermelidir (Bjorner ve diğerleri, 2007).

Soru bankası geliştirilirken, Klasik Test Teorisi (Classical Test Theory) ve Madde Tepki Kuramından (Item Response Theory) yararlanılabilir.

Aşağıda bu kuramlar ile ilgili özet bilgiler verilmiştir;

Klasik Test Teorisi (KTT) ve Madde Tepki Kuramı (MTK)

KTT’de bir bireyin bir teste ilişkin puanı; gerçek ve hata puanları toplamından oluşur. KKT’de bir bireyin bir özelliğe sahip oluş derecesi, ölçme aracının maddelerine verdiği cevapların toplamından elde edilir. Bu özelliğe sahip oluş derecesi, ölçme aracından elde edilen ham puana karşılık gelmektedir. KTT, bireylerin ölçülen özelliğe sahip oluş derecelerinin kullanılan teste bağımlı olması, testin madde istatistiklerinin de testin uygulandığı gruba bağımlı olmasından dolayı yetersiz kalmaktadır. KTT’nin sınırlılıklarını aşma çalışmaları, MTK’nın gelişmesiyle sonuçlanmıştır.

Hambleton, Swaminathan ve Roger’a (1991) göre, MTK’nın iki temel dayanağı vardır: a) bireyin testteki bir maddedeki performansı özellik, örtük özellik ya da yetenek olarak adlandırılan bir grup faktör ile açıklanabilir; b) bireyin maddedeki performansı ve madde performansının altında yatan özellikler seti arasındaki ilişki, “madde karakteristik fonksiyonu ya da madde

karakteristik eğrisi” olarak adlandırılan bir fonksiyon ile tanımlanabilir. Bu fonksiyon özellik düzeyi arttıkça, doğru yanıt verme olasılığının da artacağını gösterir. Ayrıca Klasik Test Kuramında puanlama yapılırken doğru yanıtlanan toplam madde sayısı dikkate alınırken MTK’da maddelerin ayırt edicilik ve güçlük değerleri de puanlamaya katıldığından MTK’da yetenek kestirimleri ağırlıklı puanlama ile yapılır. Bu durumda aynı sayıda doğru cevaplanan madde, öğrencilerin aynı yetenek düzeyinde oldukları anlamına gelmemektedir. MTK’ya dayalı değerlendirmelerle, ölçülen özellikle ilgili bireysel farklılıkların daha iyi ortaya konması mümkün olmaktadır (Hambleton ve diğerleri, 1991, Demirtaşlı ve Arıkan, 2009).

MTK’da, bireylerin maddelere verdikleri yanıtlara ve kullanılan modele bağlı olarak madde güçlük parametresi (b), madde ayırt edicilik parametresi (a), şansa bağlı tahmin parametresi (c) ve bireyin yetenek düzeyi (θ) tahmin edilir (Langer ve Swanson, 2010; Hambleton ve diğerleri, 1991).

Soru bankası geliştirilirken MTK’nın farklı modelleri kullanılabilir. MTK modelleri içinde Rasch modeli soru bankası geliştirilmesinde potansiyel pek çok avantaja sahiptir (Tavakol ve Dennick, 2013).

Rasch Modeli

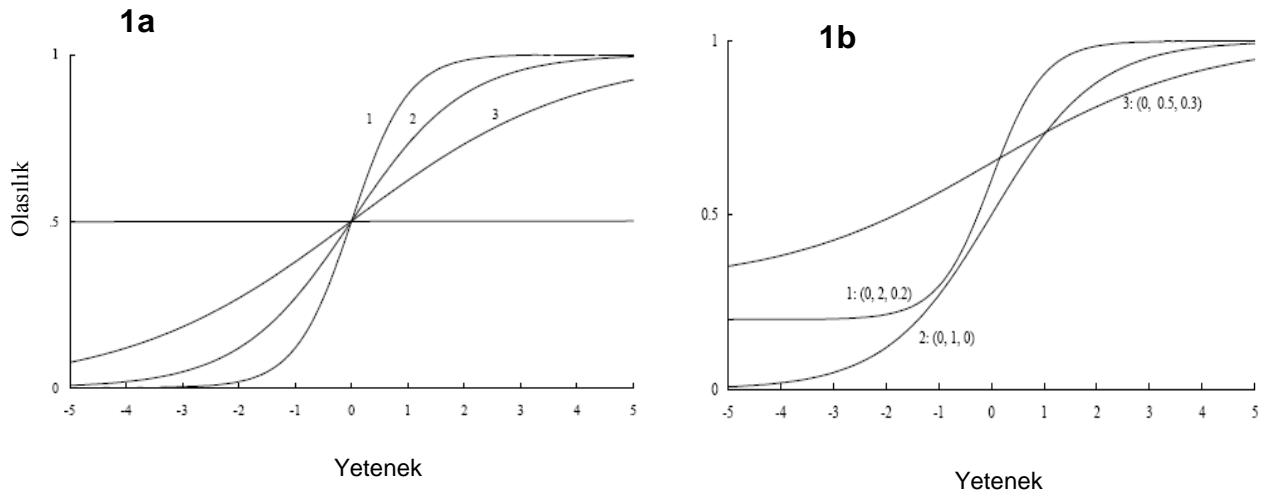
Rasch modeli, Danimarkalı matematikçi Georg Rasch tarafından 1960 yılında ortaya konulmuştur. Bu model, matematiksel olarak: madde ayırt edicilik (a_i) değerini, tüm maddeler üzerinden a değerlerinin ortalamasına eşit ve şans parametresini (c_i) sıfır kabul eden ve sadece madde güçlüğü’nün kestirildiği (b_i) bir parametrelili modelle benzer bir yapıya sahiptir. Ancak Rasch modelinde a değeri tüm modeller için 1 olarak alınır.

Rasch modelinin avantajları aşağıda açıklanmıştır;

- Ölçme kuramında;
 - İncelenen herhangi bir madde için incelenen özellik düzeyi arttıkça, o maddeye verilen doğru yanıt olasılığının da artması gerekir.

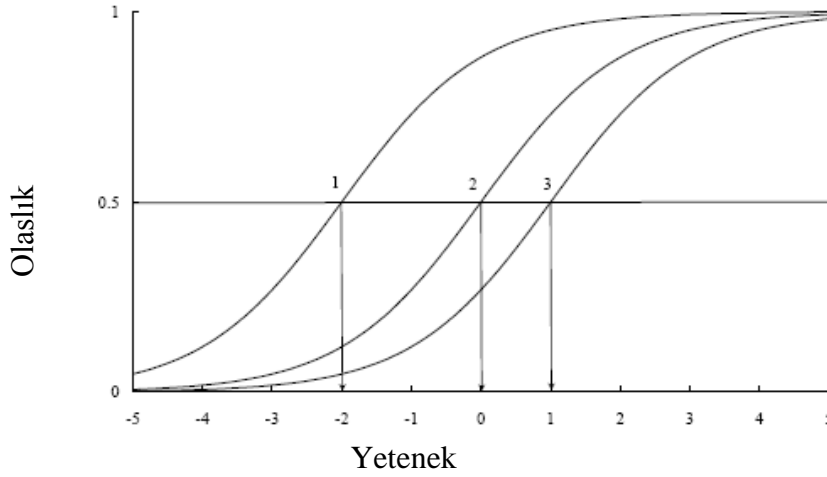
- Herhangi bir birey için, maddenin zorluk düzeyi arttıkça, maddeye doğru yanıt verme olasılığı azalır, maddenin zorluk düzeyi azaldıkça, maddeye doğru yanıt verme olasılığı artar. Diğer bir deyişle maddenin zorluk sıralaması kişiden kişiye değişmez.

İki ve üç parametrelilikte bu iki özellikten sadece birincisi, sağlanırken, 2. koşul bu modellerde sağlanamamaktadır.



Şekil 1. a) İki Parametrelilikte Üç Maddenin Karakteristik Eğrileri, **b)** Üç Parametrelilikte Üç Maddenin Karakteristik Eğrileri

İki parametrelilikte yetenek düzeyi 0'ın altında olan bireyler için madde olasılıkları $3 > 2 > 1$ olarak sıralanırken, yetenek düzeyleri 0'ın üstünde olan bireylerde madde olasılıkları $1 > 2 > 3$ olarak sıralanmaktadır ve yetenek düzeyleri 0 olan bireylerde ise tüm maddelerde olasılık eşittir (Şekil 1a). Bu durumda hangi maddenin daha zor bir madde olduğunu belirlemek güçtür. Üç parametrelilikte de iki parametrelilikte benzer şekilde yetenek düzeylerine göre madde zorlukları farklılık göstermektedir (Şekil 1b). İki ve üç parametrelilikten farklı olarak Rasch modeli bu her iki koşulu da sağlamaktadır.



Şekil 2. Rasch Modelinde Üç Maddenin Madde Karakteristik Eğrileri

Şekil 2’de görüldüğü gibi, Rasch modelinde, her madde için yetenek düzeyi ne kadar artarsa, doğru cevap verme olasılığı o kadar artmaktadır (yetenek düzeyine göre değişmemektedir) ve aynı zamanda her birey için madde ne kadar kolaysa doğru cevaplama olasılığı da o kadar artmaktadır (Chung, 2010).

- Rasch modelinin önemli bir özelliği, gözlenen ham puanların, doğrusal olmayan bir dönüşüm kullanarak θ puanlarının tahmin edilmesi için yeterli bir istatistik olmasıdır. MTK’nın diğer modellerinde, ham puan yeterli istatistik değildir (Hays, Morales ve Reise, 2000).

- Rasch modelinde kestirilecek parametre sayısının diğer modellerden az olması istatistiksel kestirim sürecindeki sıkıntıları azaltmaktadır. Ayrıca Rasch modeli ile analiz daha az sayıda denek gerektirmektedir.

- Rasch modelinde, model veri uyumu diğer MTK modellerinden farklı olarak ele alınmaktadır. Rasch modelinde, araştırmadan elde edilen verilerin, modele uyumu beklenirken ve modele uyan veriler analize dahil edilirken, diğer MTK modelleri, model veri uyumunun nasıl sağlanacağını inceler ve veriye en iyi uyan modelin kullanılmasını gerekli kılar. Oysa teorik olmayan araştırma verileri, rastgele ölçme hataları, sistematik yanlılık,

yapının iyi tanımlanamaması ve maddelerin iyi hazırlanmaması nedeniyle ideal durumdan sapmalar gösterebilir. Rasch modeli ile analiz, araştırma verileri ile model tarafından oluşturulan standart set arasındaki farkı araştırarak, veri toplama aracı veya incelenen yapının ortaya konuş biçiminin nasıl geliştirilebileceğine ilişkin bilgi sağlar. Diğer MTK modelleri, bireyin yetenek kestirimini mümkün olan en iyi şekilde yapabilmek için maddelere verilen yanıtlardaki varyansı en iyi açıklayan modeli uygularlar (Sick, 2008).

Rasch modelinin diğer MTK Modellerine göre BUT uygulamaları için soru bankası geliştirmede sağladığı diğer avantajlar aşağıda sıralanmıştır:

- Rasch analizinin veri seti, Rasch modelinin beklentilerine uyum sağladığı durumda, ölçme aracından elde edilen yanıtları, aralık ölçeğe dönüşümünü sağlayarak, bireylerin doğru bir şekilde değerlendirmesi
- Genel olarak Rasch analizinin, bireyin yetenek düzeyi ile madde zorluk değerlerini ortak bir eksen boyunca yerleştirilmesi
- Verilen bir yetenek düzeyindeki bir kişinin, gerçekleştirilmesi istenen göreve karşı ne yapabileceğinin olasılığının kestirilmesi
- Rasch modelinde, araştırmadan elde edilen verilerin, modele uyumu beklenir ve modele uyan veriler analize dahil edilmesi
- Gözlenen ham puanların, doğrusal olmayan bir dönüşüm kullanılarak θ puanlarının tahmin edilmesi için yeterli bir istatistik olmasıdır (Elhan 2002; Hays ve diğerleri, 2000; Sick, 2008).

Gelişim Sınavı için soru bankası ile testlerin oluşturulmasında; yetenek düzeyi ve maddelerin doğru olarak cevaplanma olasılıklarını kestirmeye olanak sağlayan MTK'dan yararlanılması önerilmektedir (McHarg ve diğerleri, 2005). Oluşturulan bu soru bankaları Bilgisayar Uyarlamalı Test (BUT) uygulamalarının yapılabilmesi için gereklidir.

BUT yönteminin uygulanabilmesi için madde parametreleri bilinen, çok sayıda maddeden oluşan bir soru bankası gereklidir. θ 'nın tüm aralığında iyi ölçme yapmak için, soru bankası, θ 'nın geniş aralığında yayılım gösteren

yeterli sayıda maddeden oluşmalıdır. Soru bankasının tek boyutlu bir yapıya sahip olması, soru bankasındaki maddelerin bağımsız olarak çalışması, MTK modeline uyum göstermesi ve yaş, cinsiyet, eğitim durumu gibi faktörlerden etkilenmemesi soru bankasında bulunması beklenen önemli özelliklerdir (Embretson ve Reise, 2000).

Öğrencilerin bilgi kazanımlarının ölçülmesinde, öğrencide yorgunluk, bıkkınlık yaratmadan, zaman ve insan gücünden tasarruf etmeye olanak verecek şekilde daha az sayıda madde ile daha kısa sürede belirlenebilmesini sağlayan BUT yöntemi, uygun soru bankası oluşturulduğunda geleneksel kağıt-kalem testinden daha etkili olabilir. BUT yönteminin aşağıda sıralanan üstünlükleri göz önüne alındığında, gelişim sınavlarının bahsi geçen sınırlılıklarının giderilmesinde önemli katkılar sağlayacağı ifade edilebilir.

Bilgisayar Uyarlamalı Testin Avantajları Şunlardır (Çıkrıkcı-Demirtaşlı, 1999, Weiss ve Kingsbury,1984; Weiss, 2004; Sands, Waters ve McBride, 1997);

- Test istendiği anda uygulanır ve sonuçlar anında belirlenir.
- Testler bireysel olarak uygulandığı için bireylere, test uygulayıcılarına zamanlama için değişik seçenekler sunar.
- Bireylere teste ilişkin yazılı materyal dağıtılmadığı için, test güvenliği daha yüksektir.
- Yeterli doğruluğa (accuracy) erişmek için mümkün olan en az sayıda maddeye gereksinim vardır. Bu nedenle BUT için, sabit sayıda maddeden oluşan kağıt-kalem testine göre daha az zamana ihtiyaç duyulur. Kağıt-kalem testinde, bireyler testte yer alan tüm maddelere yanıt verirler ve uygun kesinlik derecesini elde etmek için çok fazla sayıda madde ve bireye gereksinim vardır. BUT yöntemi, aynı düzeyde güvenilirliği sağlayarak, test zamanında %50'lik bir azalmaya neden olur. Daha kısa uygulama süresi, bireyin test sonuçlarını etkileyebilecek bir faktör olan bıkkınlığı azaltır.
- BUT uygulaması kalibre edilmiş maddelerden oluşan bir soru bankası ile yapıldığından, değerlendirme alışıl原因 gelen ölçme araçlarına göre daha yüksek kesinlikte sonuç verir.

- Klasik testler, genellikle orta düzeyde yeteneğe sahip olan bireyler için daha doğru sonuçlar verirken; BUT yeteneğin geniş bir aralığında doğru sonuçlar verir.
- Bireylerin puanları test uygulamasının hemen ardından bildirilebilir.

Bilgisayar Uyarlamalı Testin Sınırlılıkları Şunlardır:

- Bireylerin bilgisayar ortamına aşinalıklarında farklılıklar olması
- MTK'ya dayalı uygulamalar nedeniyle ölçülen özelliğin tek boyutluluk (unidimensionality) varsayımının karşılanmasında yaşanabilecek güçlükler
- Geniş bir madde havuzuna gereksinim duyulması
- Bilgisayar ve özel bilgisayar programının getirdiği maliyet (Çıkrıkçı-Demirtaşlı, 1999, Weiss ve Kingsbury,1984; Weiss, 2004; Sands, Waters ve McBride, 1997).

Gelişim Sınavı uygulayan tıp fakültelerinin sayısı artmakla birlikte, Gelişim Sınavının yukarıda sayılan sınırlılıklarını azaltmaya ve Gelişim Sınavı uygulamasını en etkili, amacına en uygun hale getirecek tasarımı geliştirmeye yönelik teorik ve pratik bilgiler sunabilecek araştırmaların sayısı oldukça azdır (Ricketts, ve diğerleri, 2010).

Gelişim Sınavı'nın daha yaygın olarak kullanılması ve hazırlanma amacına çok daha uygun olarak kullanılabilmesi için gelişim sınavlarının van der Vleuten (1996a) tarafından önerilen yararlılık çerçevesi kapsamında değerlendirilmesi önemlidir. Bu çerçeveye göre bir ölçme aracı değerlendirilirken sadece psikometrik özellikleri ile değil eğitim programı ile ilişkili diğer özellikleri de göz önüne alınmalıdır.

Bu çerçeve aşağıda özetlenmiştir;

Yararlılık : *Güvenirlilik x Geçerlilik x Eğitimsel Etki x Kabul Edilebilirlik x Maliyet*

Geçerlilik: Bir ölçme aracının ölçmeyi amaçladığı özelliği, herhangi bir başka özellikle karıştırmadan, doğru olarak ölçebilme derecesidir. Yani aracın amaca hizmet etme derecesidir.

Güvenilirlik: Ölçmenin tutarlılığı ile ilgilidir. Bir testi değişik zamanlarda alan kişilerin, o testten aldıkları puanlar veya sıralamaları değişmiyorsa o test güvenilirdir.

Eğitimsel Etki: Ölçme yöntemleri, öğrencinin nasıl çalışacağını ve dolayısıyla nasıl ve ne öğreneceğini belirler (Ölçme öğrenmeyi: İçeriği; yapısı ve formatı; içerdiği sorularda verilen bilgiler; sıklığı, zamanlaması ile etkiler).

Maliyeti: Uygulanan ölçme aracının, en az zaman, emek ve maliyetle en verimli niteliği elde etmesidir.

Gelişim Sınavının yararlılık çerçevesinde değerlendirilerek irdelenmesi ve yukarıda bahsedilen kısıtlılıkların giderilmesine yönelik öneriler getirecek daha çok sayıda çalışmanın yapılmasına gereksinim vardır (Langer ve Swanson, 2010).

Türkiye’de tıp eğitiminde Gelişim Sınavının uygulanması henüz yeterince yaygınlaşmamıştır. Gelişim Sınavı, Ankara Üniversitesi Tıp Fakültesi, Ege Üniversitesi Tıp Fakültesi, Adnan Menderes Üniversitesi Tıp Fakültesi’nde uygulanmaktadır. Tıp Fakülteleri dışında, Gülhane Askeri Tıp Akademisi Hemşirelik Yüksek Okulunda da Gelişim Sınavı uygulanmaktadır. Ancak Gelişim Sınavı uygulayan fakülte ve okulların sayısı bütün içinde çok düşük bir oranı oluşturmaktadır. Gelişim Sınavının daha önce ifade edilmiş olan sınırlılıkları, Türkiye’de de sınırlı sayıda Tıp Fakültesinde yapılıyor olmasının nedenini oluşturabilir.

Eğitim, sağlık alanlarında, Madde Tepki Kuramının soru bankası geliştirmede gittikçe artan kullanımına karşın, tıp eğitiminde Madde Tepki Kuramına dayalı soru bankası geliştirme çalışmaları yaygın değildir. Gelişim Sınavı uygulamaları için kapsamlı ve nitelikli çok sayıda maddeye ihtiyaç olduğunu bildiren çalışmalara karşın, benzer şekilde Gelişim Sınavı için Madde Tepki Kuramı ile soru bankası oluşturmaya ilişkin çalışmalara rastlanmamıştır. Ayrıca Gelişim Sınavının, yüksek güvenirlik düzeyi ile bireylere ilişkin sonuçları daha kısa sürede, daha az sayıda madde ile elde edilmesini sağlayabilen Bilgisayar Uyarlamalı Test yöntemi ile yapıldığına ilişkin de alan yazında çalışmaya rastlanmamıştır.

Gelişim Sınavları için Madde Tepki Kuramı ile soru bankası oluşturulmasının ve BUT yöntemi ile Gelişim Sınavı uygulanmasının, bu

sınavın hazırlık, uygulama ve değerlendirme aşamalarında ortaya çıkan sınırlılıklarını azaltmaya yönelik katkılarını değerlendirmeye ve Gelişim Sınavı uygulamalarını geliştirmeye yönelik farklı sınav tasarımlarının oluşturulmasına ilişkin çalışmaların yapılmasına ihtiyaç olduğu görülmektedir.

Araştırmanın Amacı

Bu araştırmanın genel amacı, Ankara Üniversitesi Tıp Fakültesinde yürütülmekte olan gelişim sınavı için bir soru bankası oluşturulması ve bu soru bankası kullanılarak Bilgisayar Uyarlamalı Test (BUT) yönteminin uygulanabilirliğinin ortaya konulmasıdır.

Yukarıda yer alan amaçlar çerçevesinde aşağıdaki sorulara yanıt aranmıştır:

1. Gelişim Sınavı için tıbbi bilgi düzeyini ölçmek amacı ile geçerli ve güvenilir testlerin uygulanabileceği bir soru bankası oluşturulabilir mi?
2. İncelenen özellik düzeylerinin (θ) güvenilir olarak kestirilmesinde, Gelişim Sınavının BUT ve kâğıt-kalem testi yöntemi ile uygulanması arasında madde sayısı açısından manidar bir fark var mıdır?

Önem

Öğrencilerin bilgi kazanımlarına ilişkin maksimum bilgi sağlayan, MTK ile parametreleri kestirilmiş maddelerden oluşan bir soru bankası oluşturulması ve bireye uyarlanmış gelişim sınavı gerçekleştirilmesinin uygulanabilirliğinin ortaya konulmasını amaçlayan bu çalışmanın sonuçları, Gelişim Sınavlarının tasarlanmasında ve kısıtlılıklarının azaltılmasında yol gösterici olmasına katkı getireceği beklenmektedir. Önerilen çalışmanın hayata geçirilmesi çabalarının, gelişim sınavlarının etkin şekilde yapılması ve bu yolla tıpta ölçme değerlendirme süreçlerinin gelişimine katkıda bulunacağı düşünülmektedir.

Gelişim Sınavının hazırlama, uygulama ve değerlendirme aşamalarında ortaya çıkan kısıtlılıklarını gidermede BUT uygulamasının sağlayacağı katkılarının değerlendirilmesi önemli katkılar sağlayabilir. Literatürde Gelişim Sınavının BUT yöntemi kullanılarak uygulanmasına ilişkin örnek bulunmaması nedeniyle, çalışmanın sonuçları bu konuya ilişkin ilk örneği oluşturacaktır. Araştırma bu yönüyle de önemli bulunmaktadır.

Bu çalışmanın, ülkemizde tıp eğitimi gelişim sınavları için uygun soru bankası geliştirme ve BUT yönteminin uygulanması bakımından da bir ilk ve diğer çalışmalara referans olacağı düşünülmektedir.

Varsayımlar

Benzetim çalışmaları, BUT yöntemi kullanılarak seçilen maddelere verilen yanıtların, gerçek BUT uygulamasında elde edilecek olanlarla aynı olduğu varsayımına dayalı olarak gerçekleştirilir.

Sınırlılıklar

Gelişim Sınavı yaygın olarak tıp eğitiminde uygulanmakla birlikte, diş hekimliği, fizyoterapi, hemşirelik ve ebelik eğitimi ile mezuniyet sonrası asistan eğitiminde de uygulama örnekleri mevcuttur. Bu çalışma, gelişim sınavının mezuniyet öncesi tıp eğitiminde uygulanması ile sınırlıdır.

Tanımlar

Gelişim Sınavı: Ankara Üniversitesi Tıp Fakültesi'nde gelişim sınavı, fakültedeki tüm öğrencilerin aynı anda, aynı maddelerle karşılaştıkları, herhangi bir modül, staj ya da dönem sonuna özel olarak yapılmayan, öğrencilerin gelişim ve seviyelerini ölçmeye yönelik olarak yılda bir kez yapılan bir sınavdır. Kapsamı tıp öğrencilerinin mezun olduklarında edinmeleri gereken bilgi bütünlüğü göz önünde tutularak belirlenen her bir gelişim sınavı 200 çoktan seçmeli soru içermektedir.

Dönem, Modül: Ankara Üniversitesi Tıp Fakültesinde her ders yılı “Dönem” olarak adlandırılmaktadır ve eğitim 6 dönemden oluşmaktadır. İlk üç Dönem (Dönem I, II ve III) temel tıp bilimleri eğitimini, son üç yıl da klinik bilimler eğitimini (Dönem IV, V ve IV) kapsamaktadır. Eğitim, kuramsal dersler (sunumlar, paneller, probleme dayalı öğrenme oturumları, seminerler), uygulamalar (klinik, laboratuvar ve alan çalışmaları) ve benzeri eğitim etkinliklerinin bütünleştirilmesiyle oluşturulan modüller olarak yapılandırılmıştır. Eğitim programı, Dönem I’de üç modül, Dönem II ve III’de ise dört modül olarak oluşturulmuştur.

Probleme Dayalı Öğrenme (PDÖ): Öğrencilerin, küçük gruplar halinde bir olgu üzerinde çalışarak problemi çözerken, hangi bilgileri edinmeleri gerektiğine karar vermeleri ve bunlara ulaşmalarını temel alan bir öğrenme yöntemidir.

Ankara Üniversitesi Tıp Fakültesi Modül programında, her biri ikişer saat süren iki veya üç PDÖ oturumu yer almaktadır ve bu süre boyunca bir senaryo tartışılmaktadır. Her PDÖ grubu, on öğrenci ve bir yönlendirici öğretim üyesinden oluşmaktadır. PDÖ gruplarına öğrenci seçimi rastgele yapılmakta ve her modül (dört alt modül- sekiz hafta) sonunda PDÖ grupları yeniden düzenlenmektedir. PDÖ yönlendiricileri bir modül boyunca görevlendirilmektedir.

Kısaltmalar

PDÖ	Probleme Dayalı Öğrenme
GS	Gelişim Sınavı
KTT	Klasik Test Teorisi
MTK	Madde Tepki Kuramı
USMLE	United States Medical Licensing Examination
TUS	Tıpta Uzmanlık Sınavı
ÇEP	Çekirdek Eğitim Programı
ICD	International Classification of Disease
NBME	National Board of Medical Examiners
ACT	American Collage Test
GED	General Education Development
BUT	Bilgisayar Uyarlamalı Test
MFF	Maddenin Farklı Fonksiyonlaşma

BÖLÜM II

KURAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR

Bu bölümde Gelişim Sınavı, Bilgisayar Uyarlamalı Testler, soru bankası ile ilgili kuramsal bilgi ve araştırmalar sunulmaktadır.

Gelişim Sınavı

1970’li yıllarda Kanada McMaster Üniversitesi ve hemen ardından da Hollanda Maastrich Tıp Fakültesinde Probleme Dayalı Öğrenme uygulamalarının başlaması ile birlikte yeni değerlendirme yöntemlerine ihtiyaç duyulmuştur. PDÖ ile öğrenci merkezli, derinlemesine ve hayat boyu öğrenme temelinde bilgi kazanımlarının, hatırlama ve test yönelimli öğrenmeyi teşvik eden klasik çoktan seçmeli sınavlarla değerlendirilemeyeceği görüşünden yola çıkılarak 1970’li yılların sonunda hem Maastric Üniversitesi hem de Missouri Üniversitesinde Gelişim Sınavı uygulanmaya başlamıştır. O günden bu yana da dünya genelinde Tıp Eğitimi programlarında yaygın olarak kullanılmaya başlanmıştır. Çalışmalar bu boylamsal ve daha çok çoktan seçmeli testlerle yapılan sınavın Antarktika dışında, Güney Afrika, Asya, Avrupa, Orta Doğu, Doğu, Güney ve Kuzey Amerika, Yeni Zellanda ve Avusturalya’yı kapsayacak şekilde hemen hemen bütün kıta/bölgelerde uygulanmakta olduğunu göstermektedir (Freeman ve diğerleri, 2010) (Ek1).

Gelişim Sınavı Uygulamaları

Gelişim Sınavı İngiltere (Swanson ve diğerleri, 2010), Hollanda (Schuwirth, Bosman, Henning, Rinkel ve Wenink, 2010) ve Almanya’da (Avustruya da dahil olmak üzere) (Nouns ve Georg 2010) ulusal

konsorsiyumla/ortaklığı; Afrika (Aarts, Steidel, Manue, ve Driessen, 2010), Suudi Arabistan (Al Alwan ve diğerleri, 2011), Güney Doğu Asya, Karayipler, Avustralya, Yeni Zellenda, İsveç, Fillandiya, İngiltere ve Amerika Birleşik Devletlerindeki Tıp Fakültelerinde kurumsal olarak yapılmaktadır (Freeman ve diğerleri, 2010). Amerikada bulunan Ulusal Tıp Sınavları Kurulu (National Board of Medical Examiners- NBME) farklı ülkelerde gelişim sınavlarının yürütücülüğünü yapmaktadır (De Champlain ve diğerleri, 2010; Uluslararası Tıp Kurum- International Foundations of Medicine, 2011). Kanada'da kurulmuş olan uluslararası konsorsiyuma; İrlanda, Avustralya, Kanada, Portekiz ve Batı Hint adalarından okullar katılmıştır (Finucane, Flannery, Keane ve Norman, 2010; International Partnership for Progress Testing, 2011).

Tıp eğitiminde kullanılan bir değerlendirme aracı olarak Gelişim Sınavı diğer değerlendirme yöntem ve araçlarından farklı özellikler göstermektedir. Tıp eğitimi programı süresince, programdaki tüm öğrencilere aynı zamanda ve belirli aralıklarla uygulanmaktadır.

Gelişim Sınavında Kullanılan Madde Türleri ve Madde Sayıları

Gelişim sınavları geniş bir konu alanında, geniş bir madde örneklemini gerektirdiğinden genel olarak çoktan seçmeli sorular kullanılarak hazırlanmaktadır. 1977 yılından bu yana gelişim sınavı uygulamakta olan Maastrich Üniversitesi Tıp Fakültesinde gelişim sınavı 250 doğru yanlış sorusundan oluşmaktadır. Missouri Kansas City Tıp Okulunda (ABD) yürütülmekte olan gelişim sınavı (the Quarterly Profile Examination-QPE) 400 çoktan seçmeli soru, Mc Master Üniversitesinde (Kanada) yürütülmekte olan ise (Personel Profile Index-PPI) 180 çoktan seçmeli soru içermektedir. Bunun yanı sıra Dundee Tıp Fakültesi gelişim sınavı yapılandırılmış cevaplı sorular (CRQ's) ile, Utrecht Üniversitesi Tıp Okulunda (Hollanda) gelişim sınavı açık uçlu kısa cevaplı sorular ile yürütülmektedir (Freeman ve diğerleri, 2010; Rademakers ve diğerleri, 2005). EK1 'de yer alan Çizelge 6'da uluslararası alanda çeşitli tıp fakülterinin gelişim sınavı uygulamaları, madde formları ve madde sayıları gösterilmiştir.

Öğrenci puanları, boylamsal (longitudinal) olarak ve tekrarlı ölçmelerle elde edilmekte, bu puanlar tıp eğitiminin bütününe ilişkin hedeflerin müfredattan bağımsız olarak değerlendirilmesine olanak sağlamaktadır (van der Vleuten ve diğerleri, 1996). Bu özellikleri Gelişim Sınavının tıp eğitiminde önemli bir yeri olmasını sağlamaktadır. Hollanda, Kanada, İngiltere, İrlanda'da tıp fakültelerinde mezuniyet öncesi tıp eğitimine ilişkin yapılan çalışmalar Gelişim Sınavının boylamsal olma özelliğinin, eğitim süresince öğrencilerin bilgi kazanmalarının gelişimi ve etkililiğine yönelik gözlenebilir ve özgün değerlendirmeler sağladığını göstermektedir (Verhoeven ve diğerleri, 2002. Coombes, Ricketts, Freeman ve Stratford, 2010; Finucane ve diğerleri, 2010; Freeman ve diğerleri, 2010; Schaap, Schmidt ve Verkoeijen, 2012).

Uygulanışı ve Puanlamasındaki Farklılıklar

Kabul edilen önemi, avantajlarının ve giderek artan kullanımının yanı sıra Gelişim Sınavının uygulanışı ve içeriğine ilişkin önemli farklılıklar görülmektedir (Ricketts ve diğerleri, 2010). Belirtke tablosu ve içerik örnekleme profili, madde sayısı ve uygulanma sıklığı, “şans başarısı düzeltme” formülü kullanma durumu uygulayıcılara göre değişiklik göstermektedir (McHarg ve diğerleri, 2005). Ayrıca, testin uygulanış amacı (biçimlendirmeye yönelik, belgelendirmeye yönelik) ve puanlama yöntemleri uygulayıcılara göre farklılaşmaktadır.

Gelişim Sınavının planlaması ve uygulanması süreci

Bu süreç 4 ana başlık içermektedir (Wrigley, van der Vleuten, Freeman, Muijtjens, 2012):

1. Sınavı yapılandırma:

Belirtke tablosu/kapsam matrisi: Gelişim Sınavı hazırlama sürecinde geçerliği ve güvenilirliği sağlama açısından temel bileşeni oluşturur. Belirtke tablosu kapsam geçerliğini ve böylece aynı içeriği temsil edecek şekilde

madde hazırlanması yolu ile karşılaştırılabilirliği sağlar (Bridge, Musial, Frank, Roe ve Sawilowsky, 2003; Muijtjend ve Wijnen, 2010; Swanson ve diğerleri, 2010). Gelişim Sınavı için yaygın kullanılan belirtke tablosu/kapsam matrisi genellikle kolonlarını organ sistemleri (solunum sistemi, kas-iskelete sistemi gibi) ve satırlarını disiplinler (Anatomi, Genel Cerrahi) ya da işlem ve görevleri (task) (hastalık mekanizmaları, tedavi prensipleri vb) içeren bir tablo olarak hazırlanmaktadır (Coombes ve diğerleri, 2010; Muijtjens ve Wijnen 2010; Nouns ve Georg 2010; Swanson ve diğerleri, 2010). Belirtke tablosu/matristeki her bir hücre, eğitim programının nihai bilgi hedeflerinin önemli olma derecesi ya da önceliğine göre ağırlıklandırılır ve böylece her test için bu hücreden hazırlanacak madde sayısı belirlenir.

Bazı kurumların Gelişim Sınavı için hazırlamış oldukları belirtke tablolarında zihinsel süreçlerin farklı düzeyleri de dikkate alınmaktadır. Bu bağlamda belirtke tablosunun hazırlanması iki ardışık karar sürecini kapsar. Bunlardan birincisi, matrisin kapsamı, ikincisi ise matrisin her bir hücresine ilişkin ağırlıklandırılmış madde sayısını belirleme oluşturmaktadır.

Belirtke tablosunun içeriği programın sonunda kazanılması beklenen program yeterliklerine bağlı olarak ve bu yeterlikleri yansıtabilecek şekilde oluşturulmaktadır. Belirtke tablosunun doğrudan müfredata bağlı olmaması, ulusal ve kurumsal düzeyde belirlenmiş program yeterliklerini temel alması Gelişim Sınavının önemli bir özelliğini oluşturmaktadır.

Madde yazımı: Gelişim Sınavı için madde yazımının temel zorluğunu, maddelerin belirli bir ünite veya dersle doğrudan ilişkili olarak değil ancak daha kapsamlı olarak hazırlanmış olan program yeterliklerine bağlı olarak hazırlanması oluşturmaktadır. Alan yazında, Gelişim Sınavı için hazırlanan maddelerin, tek doğru cevaplı, birden çok doğru cevaplı (alternative-choice) doğru-yanlış, çoktan seçmeli ve eşleştirmeli madde formatlarında olduğu belirtilmektedir. Yine literatürde/alan yazında doğru-yanlış formatında hazırlanan maddelerin güvenilirliğinin düşük olduğu (Zimmerman ve Williams, 2003), bununla birlikte tek doğru cevaplı çoktan seçmeli test maddelerinin şansa doğru cevaplanma olasılıklarının daha düşük olduğu ve daha güvenilir sonuçlar verdiği belirtilmektedir. Farklı ülkelerde ve üniversitelerde şans

başarısını azaltmaya yönelik olarak “bilmiyorum” seçeneğinin eklenmesi ve tahminde bulunmanın azaltılması için şans başarısı düzeltme formülü kullanılmaktadır. Zimmerman ve Williams (2003) yaptıkları çalışmalarında, çoktan seçmeli sınavlarda tahminden kaynaklanan hatanın diğer varyans kaynaklarından daha fazla olduğunu göstermişlerdir. Wade ve diğerlerinin (2011), tahmin etmeyi kontrol altına almaya yönelik uygulamalara ilişkin yapmış oldukları çalışmalarında, öğrencilerin doğru cevabı bulmada, şans ve tahminin başarılarının, bilgilerinden daha çok yardımcı olduğuna inandıklarını bulmuşlardır. Bu görüşün şans başarısı için düzeltme formülü uygulamayan fakültelerde daha yaygın olduğu saptanmıştır.

Soru Bankası: Gelişim Sınavı uygulamaları için güvenliği sağlanmış soru bankasına ihtiyaç vardır. Soru bankasının büyüklüğünü, her yıl yapılan Gelişim Sınavı sayısı, her sınavdaki madde sayısı, maddelerin tekrar kullanılmasına ilişkin politikalar, soruların ve cevapların öğrencilerle paylaşılıp paylaşılmaması etkilemektedir.

Değerlendirme komisyonu: Maddelerin ardışık süreçlerle kontrol edilmesi ve gözden geçirilmesi, Gelişim Sınavının kalitesinin geliştirilmesi için son derece önemli bir süreçtir. Gelişim Sınavı uygulayanlar için önerilen, maddelerin hazırlanması, uygunluğunun kontrolü, testin yapılandırılması ve yönetim süreçlerinden sorumlu bir ekibin kurumsal düzeyde oluşturulmasıdır.

II. Testi uygulama:

Testin amacı: Gelişim Sınavı biçimlendirmeye veya belgelendirmeye yönelik olabilmektedir. Sınavın kullanımı bu açıdan kurumdan kuruma farklılık göstermektedir. Örneğin, Hollanda konsorsiyumunda Gelişim Sınavı sonuçları belgelendirmeye yönelik değerlendirme amacıyla kullanılmaktadır. Öğrencilerin bir sonraki eğitim aşamasına geçebilmeleri için Gelişim Sınavı'ndan geçer not almaları gerekmektedir. Belgelendirmeye yönelik kullanımın yanı sıra öğrenciye kapsamlı geribildirim de sağlanmaktadır. Ancak, Almanya ve Kanada konsorsiyumlarında Gelişim Sınavı

biçimlendirmeye yönelik değerlendirme amacıyla yapılmakta ve öğrencilerin öğrenmelerini geliştirmeye yönelik kapsamlı geribildirimler sağlanmaktadır.

Gelişim Sınavı'nın eşzamanlılığı: kurumda ya da konsorsiyumlarda yapılan Gelişim Sınavlarında, tüm kurumlardaki tüm öğrencilere aynı anda sınav yapılmaktadır. Sınavın eş zamanlı olarak uygulanması lojistik ve kullanılan kaynaklar açısından güçlükler yaratmaktadır. Öğrencilerin sınavı farklı zamanlarda alması da, aynı psikometrik özelliklere sahip farklı sınav formlarına gereksinim doğurmaktadır. Bu da çok daha büyük bir madde havuzu ve madde yazımı için daha fazla insan gücü ve zaman gerektirmektedir.

BUT yöntemi eşzamanlılıktan kaynaklanan zorlukları azaltabilir. Ancak tüm maddelerin önceden denenmiş ve kalibrasyonlarının yapılmış, parametrelerinin tanımlanmış olması gibi bir ön koşulu olduğu da göz önünde bulundurulmalıdır.

Dönemler, test sıklığı ve test büyüklüğü: Bazı farklılıklar olmakla birlikte dünyada Gelişim Sınavının eğitim programının tüm aşamalarındaki öğrencilere aynı zamanda uygulanmasının yaygın olduğu görülmektedir (Freeman ve diğerleri, 2010). Testin uygulanma sıklığı ve madde sayısına ilişkin kabul görmüş kılavuzlar bulunmamaktadır. Maliyet ve ulaşılabilir kaynaklar uygulama sıklığı ve madde sayısı üzerine etkili olmakla birlikte güvenilirlik hesapları, test sıklığı ve içereceği madde sayısı için önemli bir belirleyicidir. Maastrich Üniversitesinde yapılan genellenebilirlik çalışmasında, test sıklığı ve madde sayısı arttıkça güvenilirliğin arttığı; toplam madde sayısı sabit tutulduğunda, test sıklığını arttırmanın, güvenilirliği, her test için madde sayısını arttırmaktan daha olumlu yönde etkilediği gösterilmiştir (Wrigley ve diğerleri, 2012).

Sınav süresi: Sınav süresi değişik kurumlardaki uygulamalarda farklılıklar göstermektedir, örneğin, İngiltere'de yapılan uygulamalarda sınav için 2,5 saat süre ayrılırken, NBME beş saat süre ayırmaktadır. Süre doğal olarak madde sayısı ve maddeleri okumak için gereken süre ile orantılı olarak belirlenmektedir. Ancak Hollanda konsorsiyumu öğrencilerin tahmin

etmelerine zaman yaratmayacak ancak Gelişim Sınavının hız testi olmama özelliğine zarar vermeyecek bir süre planlamanın uygun olacağını belirtmektedir.

III. Sonuçların analizi ve değerlendirme:

Bu aşama, öğrencilerin gelişimlerine ilişkin güvenilir ve savunulabilir bir değerlendirme yapmak üzere uzmanlar tarafından dikkatle planlanmalıdır.

Puan hesaplama yöntemleri: Yaygın olarak doğrudan doğru sayısı üzerinden puanlama yapma ve düzeltme formülü kullanarak toplam puanı hesaplama olmak üzere iki puanlama sistemi kullanılmaktadır. Düzeltme formülü, sadece doğruların toplanarak toplam puan elde edilirken dikkate alınmayan, tahminden kaynaklanan ölçme hatasını ortadan kaldırmak amacı ile kullanılmaktadır. Öğrencileri tahminde bulunmaktan vazgeçirmek üzere Hollanda konsorsiyumu/ortaklığı yanlış cevaplara yönelik ceza uygulamakta ve düzeltme formülü kullanılmakta, seçenek sayısına bağlı olarak belli oranlarda (örn: beş seçenekli maddeler için -0,25) yanlış cevap sayısını doğru cevap sayısından düşürmektedir. Yapılan tartışmalarda düzeltme formülünün kullanılmasının Gelişim Sınavı altında yatan felsefe ile tutarlı olduğu savunulmaktadır. Öğrencilerin bilmiyorum olarak cevapladıkları ya da boş bıraktıkları maddeler öğrencilere ve eğitimcilere bilgi kazanımındaki eksik yönleri göstermede doğrudan bilgi sağlamaktadır. Öğrencinin “bilmiyorum” olarak verdikleri cevaplar “tahmin etmeye” tercih edilmektedir Bu özellikle eğitimin ilk yıllarındaki öğrenciler için çok güç maddeleri de içeren boylamsal testler için uygundur. Tahminden kaynaklanan hata varyansı diğer olası kaynakların hata varyansından çok daha fazla bulunmuştur (Muijtsen, van Mameren, Hoogenboom, Evers ve van der Vleuten, 1999). Bu nedenle puan hesaplama yöntemini belirlemek kolay bir seçim değildir. Gerçekte puanların oluşturulmasında yaşanan bu problem, maddelerin öğrencilerin yetenek düzeyine göre belirlenebildiği BUT yönteminin kullanılması ile giderilebilir (Roex ve Degryse, 2004). Bu yaklaşımla “tahmin” bir sorun olmaktan çıkmakta ve “bilmiyorum” seçeneğine gerek kalmamaktadır.

Alan yazında tıp eğitiminde Gelişim Sınavı uygulamalarında Bilgisayar Uyarlamalı Test (BUT) uygulamalarına ilişkin bir yayın bulunmamaktadır (Roex ve Degryse, 2004).

Standart oluşturma yöntemleri: Literatürde/alan yazında farklı standart oluşturma yöntemleri sıkça tartışılmaktadır. Yaygın olarak kullanılan yöntemler; norm dayanaklı ve kriter dayanaklı puanlama yöntemleridir ve bu konuda hala tartışmalar sürmektedir. Her iki yöntem de avantaj ve dezavantajlara sahiptir (Muijtjens, Hoogenboom, Verwijnen ve van der Vleuten,1998). Gelişim Sınavlarının güçlük düzeylerinin farklılık gösterebileceği gözönüne alındığında mutlak, sabit bir kesme puanının kullanılmasının norm dayanaklı puanlardan daha riskli olduğu savunulmaktadır. Hollanda ve Almanya konsorsiyumunda puanlar norm dayanaklı olarak hesaplanmaktadır (Muijtjens ve diğerleri, 2008; Schaubert ve Nouns, 2010).

Yerel/ulusal değerlendirme komitesi: Gelişim Sınavının kapsam geçerliğini ve güvenilirliğini arttırmada önemli bir bileşen test maddelerinin sınav öncesi ve sonrası uzmanlar tarafından değerlendirilmesidir. Sınav sonrası değerlendirmede uzmanlardan, öğrencilerin puanlarının hesaplanması aşamasında hangi maddelerin dahil edilip, hangi maddelerin dahil edilmeyeceğine karar vermeleri beklenmektedir.

IV. Geribildirim:

Gelişim Sınavının altında yatan prensip, öğrencilerin derinlemesine öğrenmelerini sağlamaya yönelik olarak boylamsal ve gelişime yönelik geribildirimler sağlamaktır. Bilişsel davranışların tekrarlı olarak sınanmasını ve kapsamlı geribildirimler sağlayan Gelişim Sınavı, öğrencilere kendi öğrenmelerini yönlendirme fırsatı sağlar ve bu yolla öğrenmeyi artırır (Norman, Neville, Blake ve Mueller, 2010; Ricketts ve diğerleri, 2010). Bu bulgular, bilgiyi tekrarlama ve düzenleme fırsatı yaratan tekrarlı ölçümlerin, öğrenmeyi, akılda kalıcılığı ve bilgiyi transfer edebilmeyi arttırabileceğini gösteren araştırmalar ile uyumludur (Roediger ve Karpicke, 2006; Roediger

ve Butler, 2010). Gelişim Sınavı puanları, aynı zamanda sınavı hazırlayanlar, eğitimciler ve programı geliştirenlere de önemli bir bilgi kaynağı sağlar.

Uygun maddelerden oluşan güvenilir, geçerli Gelişim Sınavları hazırlamak için, iyi hazırlanmış, yeterli sayıda madden oluşan bir soru bankasının olması önemlidir (Ricketts ve diğerleri, 2010).

Soru Bankası

Soru bankası geliştirme, yeterli sayıda sorunun toplanmasından, bu soruların psikometrik özelliklerinin saptanması ve kalibrasyonlarının yapılmasına kadar pek çok aşamayı içermektedir.

Soru bankası oluşturmada aşağıda özetlenen iki yaklaşım kullanılabilir.

-Yukarıdan aşağıya yaklaşımı: Bu yaklaşımda büyük bir soru bankasından bir seri faktör analizi yapılarak tanımlı madde kümeleri oluşturulur. Oluşturulan bu soru kümelerinden tek boyutluluğu da sağlayacak şekilde MTK analizleri ile alt soru bankaları elde edilir.

-Aşağıdan yukarıya yaklaşım: Bu yaklaşımda, daha az ancak daha iyi tanımlanmış sorularla banka oluşturmaya başlanır (örn; depresyona ilişkin) ve daha sonra alt boyutlar bir araya getirilerek temel boyutlar oluşturulur (Chang, 2004).

Soru bankası geliştirmede temel adımlar:

- Testin kapsamında bulunan her içerik alt alanı için yeterli miktarda soru seçilmesi/soru yazılması,
- Soruların bilimsel ve teknik açıdan değerlendirilmesi,
- İlk kez kullanılacak olan sorular için ön uygulama yapılması,
- Ön uygulamalardan elde edilen madde istatistikleri (KKT ve MTK'na dayalı olarak) dikkate alınarak sorulardan alt testler oluşturulması,
- Testin kapsam geçerliliğini geliştirmeye yönelik soruların kapsam içindeki dağılımlarının düzenlenmesidir.

Soruların seçilmesi ya da yazılmasına yol göstermek üzere ölçülecek özelliğin kapsamı ve alt alanlarının belirlenmesi gerekmektedir.

BUT ile bireylere soru bankasından seçilen soruların oluşturduğu farklı test formları uygulanmaktadır, ancak bu uygulamalarda kullanılan soru bankaları, her bir maddenin her bir birey için kullanılabilme potansiyeli nedeni ile tek bir test formu gibi düşünülebilir. Bu nedenle soru bankası tüm boyutları, farklı yetenek düzeyleri için sorgulayabilecek kapsamda soru içermelidir.

Eğer ölçülecek yapı çok boyutlu ise, soru bankası da bu boyutları sorgulayabilecek soruları içermelidir. Ancak buradaki temel tartışmayı BUT uygulamaları için hazırlanan bankaların tek boyutlu olması varsayımı oluşturmaktadır. Soru bankasının içerik dengesini gözetmek bir çözüm olmakla birlikte, uyarlanmış soru seçiminin işlevliliğini etkili bir şekilde yerine getirebilmek üzere her bir boyut için bir soru bankası oluşturulması da uygun olacaktır (Wainer ve diğerleri, 2000).

BUT uygulamaları için, soru bankasının oluşturulmasında izlenecek adımlar aşağıda sıralanmaktadır (Şekil 3) (Bjorner ve diğerleri, 2007):

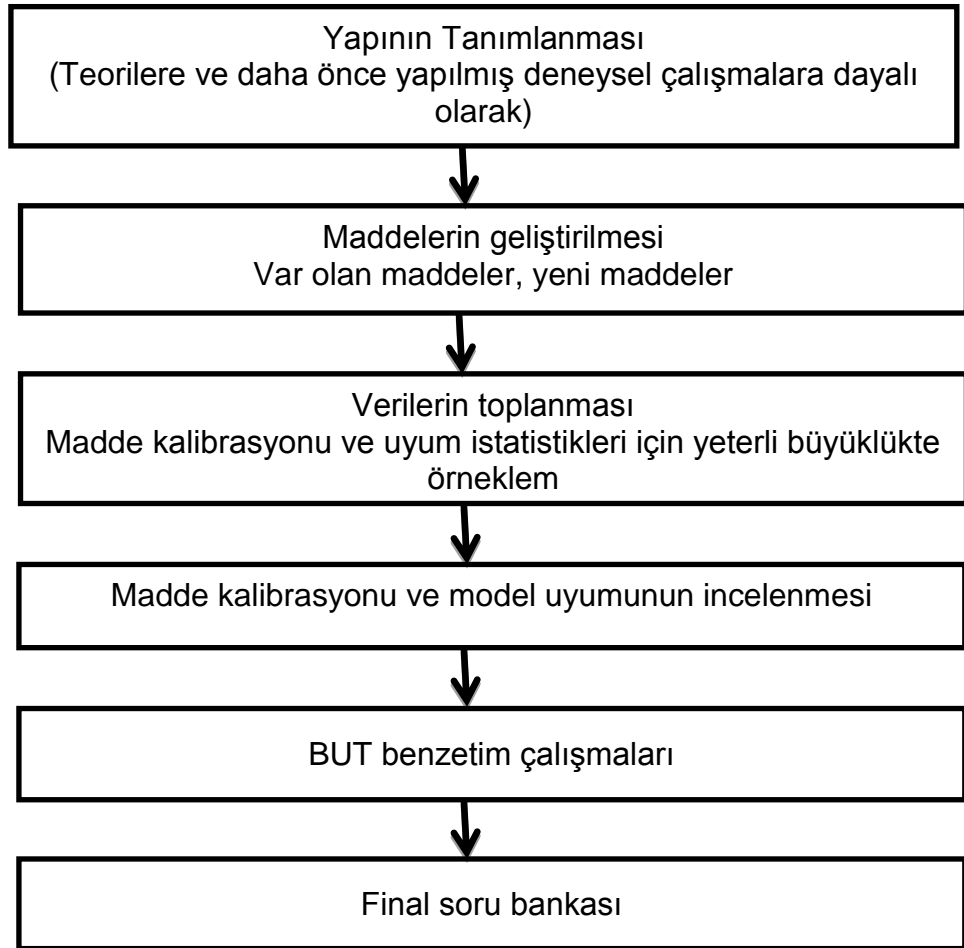
- *Ölçülecek yapının tanımlanması ve soru geliştirme:* Ölçülecek yapının, yapıya ilişkin alt alanları ile belirlenerek, soru bankasının yapının ilgili tüm özelliklerini içermesi sağlanmalıdır. Bu aşamada, daha önceden var olan sorular kullanılabileceği gibi yeni sorular da hazırlanabilir. Ancak soruların hepsinin aynı yapıyı/alt alanları ölçüyor olmasına dikkat edilmelidir.

- *Verilerin toplanması:* Modelle ilişkili parametrelerin kestirimi ve madde kalibrasyonu için yeterli örnek büyüklüğüne ihtiyaç vardır. Kullanılacak MTK modeline göre örneklem büyüklüğü farklılık göstermektedir. Bu modeller içinde en küçük örnekleme Rasch modeli ihtiyaç göstermektedir.

- *Madde kalibrasyonun yapılması ve model uyumu ile model varsayımlarının test edilmesi:* Bu aşama MTK modelinin belirlenmesi, farklı gruplardan elde edilmiş madde parametrelerinin kalibrasyonlarının yapılarak

aynı ölçek üzerinde yerlerinin belirlenmesi, belirlenen modelin varsayımlarının karşılanma durumu (tek boyutluluk, yerel bağımsızlık) modele uyum, Maddenin Farklı Fonksiyonlaşması (MFF) varlığının saptanması ve madde parametrelerinin belirlenmesi çalışmalarını içerir.

- *BUT benzetim çalışmasının yapılması*: Bu aşama soru seçimi ve durdurma kurallarının belirlenmesi, belirlenmiş olan kuralların, testin uzunluğu, güvenirliği ve geçerliği üzerine etkisinin simule BUT uygulamaları ile değerlendirilmesini içerir.



Şekil 3. Soru Bankası Geliştirme Basamakları

Farklı yeterlik düzeylerindeki öğrencilerin değerlendirilmesine olanak sağlayacak şekilde çok sayıda maddenin sağlanması ve farklı uygulamalardan elde edilen madde parametrelerinin kalibrasyonlarının

yapılarak aynı ölçek üzerinde yerlerinin belirlenmesi soru bankası geliştirilmenin kritik aşamasını oluşturmaktadır. Bu aşamada yapılacak çalışmalar şu şekilde özetlenebilir:

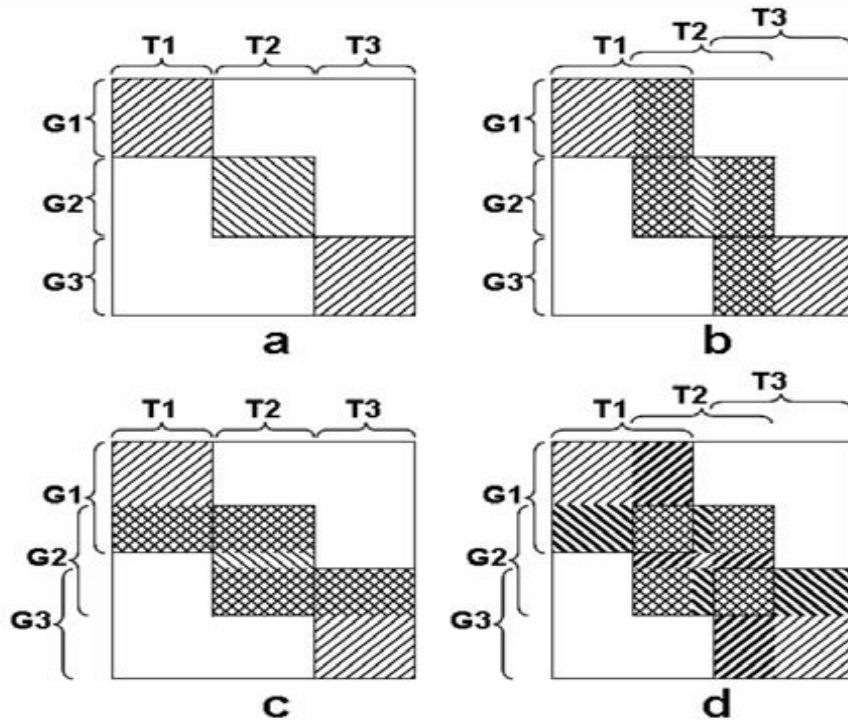
- Kalibrasyon çalışmaları için soru bankasından alt testlerin oluşturulması (Soru bankasının kalibrasyonun yapılmadan önce, soru bankası alt testleri bilgisayar ortamında ya da kağıt-kalem testi olarak uygulanmış olmalıdır).
- Soru bankası alt testlerinin uygulanması ile maddeye ilişkin cevapların elde edilmesi
- Soru bankasının kalibrasyonu için kalibrasyon veri setinin, soru bankası veri seti ile birleştirilmesi
- Soru bankasının psikometrik özelliklerini belirlemeye yönelik olarak kalibrasyon veri setinin psikometrik analizlerinin yapılması (Molina, 2008).

Kalibrasyon çalışmalarında; soru bankasındaki tüm maddelerin bireyler tarafından cevaplanması ve toplanan bu verilerin kalibrasyon çalışmalarında kullanılması uygulama açısından güçlükler içermektedir. Bu nedenle, kalibrasyon çalışmalarında yaygın kullanılan temel yöntemlerden biri; soru bankasının kabul edilebilir büyüklükte alt testlere bölünmesi ve bireylere uygulanarak kalibrasyon için gerekli olan verilerin elde edilmesidir. Soru bankasının alt testlere ayrılarak tüm maddelerin ortak bir ölçek üzerine yerleştirilmesi ve kalibrasyon çalışmaları için geliştirilmiş stratejiler kullanılabilir.

Şekil 4'de soru bankasının iki (üç) alt test ve öğrencilerin iki (üç) alt gruba bölündüğü dört kalibrasyon modeline ilişkin özet şemalar sunulmuştur (Molina, 2008).

- Şekil 4(a)'da soru bankasının birbirini kapsamayan alt testlerinin, yine birbirini kapsamayan öğrenci guruplarına uygulandığı kalibrasyon modeli gösterilmektedir. Madde parametrelerinin aynı ölçek üzerinde ilişkilendirilmesi işlemi ya testlerin ya da öğrencilerin eşitlenmesi yani denk hale getirilmesi ile gerçekleştirilmektedir.

- Şekil 4(b) anchor-madde modelini göstermektedir. Bu modelde maddelerden bazıları tüm öğrenci alt grupları tarafından cevaplanır; bu anchor maddeler, farklı öğrenci gruplarından elde edilmiş madde parametrelerini aynı ölçek üzerinde ilişkilendirmek için anahtar görevi görmektedirler.
- Şekil 4(c)'de anchor-öğrenci modeli şematize edilmiştir. Bu modelde alt gruplardaki bazı öğrenciler soru bankasındaki tüm maddeleri cevaplar ve bu cevaplar madde parametrelerini aynı ölçek üzerinde ilişkilendirmek için anahtar görevi görür. Anchor-madde modelindeki gibi, bu model madde/öğrenci alt gruplarının eşdeğer olması varsayımını gerektirmez. Ancak anchor-birey modelinde, anchor grubunu oluşturan öğrenciler tüm maddeleri cevaplamaktan dolayı yorgunluk/bıkkınlık hissedebilirler. Bu durum bu model için bir kısıtlılık yaratabilir.
- Şekil 4(d)'de ortak maddeler ve ortak öğrenci gruplarının kullanıldığı karma bir model olan çifte anchor modeli gösterilmektedir.



Şekil 4. Soru Bankasındaki Soru Kalibrasyonu için Oluşturulmuş Alt Test/Alt Grupları İçeren Modeller.

Madde parametrelerinin kalibrasyonu için, tanımlanan bu modellerin yanı sıra pek çok farklı anchor modeli bulunmaktadır. Bu kapsamda madde parametresi kestiriminde MTK'yı temel alan programlarla da (örneğin, RUMM 2020, LOGIST, BILOG vb.), bireylerin yetenek kestirimleri ve madde parametreleri için kalibrasyon çalışmaları yapmak mümkün olmaktadır (Molina, 2008).

Gelişim Sınavında BUT uygulamalarının yapılabilmesi için yukarıda özellikleri ve oluşturulma süreçleri tanımlanan soru bankaları gereklidir.

Bilgisayar Uyarlamalı Testler

“Uyarlamalı test” yöntemi, bir ölçme aracını oluşturan maddelerin uygulama sırasında seçildiği bir süreci tanımlamak için kullanılan çok sayıda terimden birisidir (Ware, 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 veya maddeler her bir birey için farklılık gösterdiğinde, “uyarlamalı” olarak tanımlanır (Blais ve Raiche, 2002). Bu kapsamda kullanılan diğer terimler; bireye göre ayarlanan test yöntemi (tailored testing), bölümlendirilmiş test yöntemi (branched testing), programlanmış test yöntemi (programmed testing), bireyselleştirilmiş test yöntemidir (individualized testing).

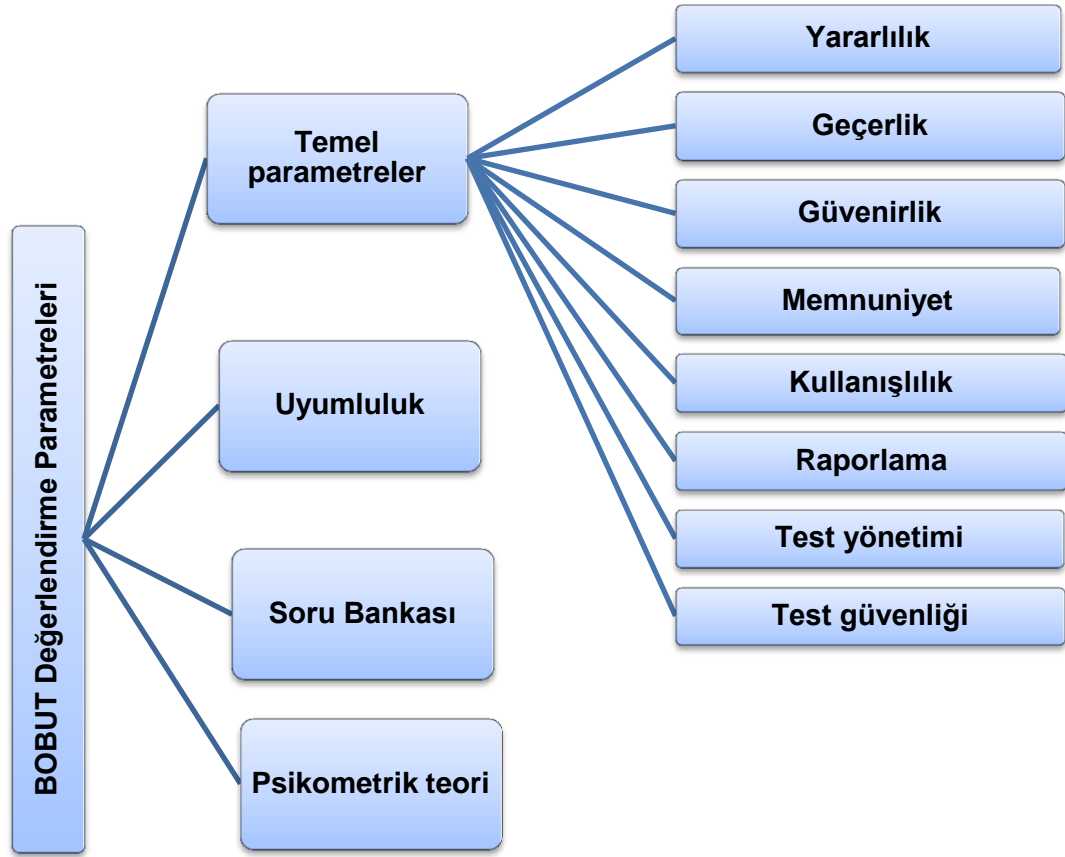
Bireye Uyarlanmış Testler (BUT), döngüsel (iteratif) bir süreç olup; maddelerin bireye uyarlanma işi ardışık olarak yapılır. Maddelerin cevaplayıcılara verilmiş sırası yanıtlayıcının önceki maddedeki performansına göre düzenlenir. Bu süreç boyunca testi alan bireylerin yetenek düzeyi sürekli olarak değerlendirilir. İlk olarak, soru bankasındaki maddelerden orta derecede zorluğa sahip olan bir madde seçilir ve birey yanıt verir. Verilen yanıtla göre bireyin θ düzeyine ilişkin tahmin elde edilir ve ardından soru bankasındaki bireyin θ tahmini için en fazla bilgi sağlayacak madde sorulur.

Maddelere verilen yanıtlara dayalı olarak, yeni bir θ tahmini hesaplanır. Bireyin θ düzeyi istenen doğruluk düzeyinde belirlendiğinde (örneğin, belirlenmiş bir standart hata miktarı) test sona erer (Roex ve Degryse, 2004; Öztuna, 2008).

BUT ilk kez Alfred Binet tarafından 1905’de geliştirilmiştir. Lord’un, 1960’ların sonunda "Eğitimsel Test Servisi"nde (Educational Testing Service – ETS, ABD) başlattığı çalışmalar BUT literatürüne/alan yazınına önemli katkılar sağlamıştır. Uygulamada test, testi alan her bireyin ölçülen yetenek düzeyine uyarlanmakta, test maddelerinin farklı setleri bireyin ölçülen özellikteki yerine bağlı olarak uygulanmaktadır. Bu durumda her birey güçlük düzeyi kendi yetenek düzeyine göre uyarlanmış farklı testleri alır.

1970’lerin başından itibaren, bilgisayar teknolojisindeki gelişmeler ve bilgisayara ulaşma ve kullanımındaki artış, MTK analizi ve uyarlanmış testler yapabilen yazılımların geliştirilmesi ile Bireye Uyarlanmış Testler "Bilgisayar Uyarlamalı Test (Computerized Adaptive Testing - CAT)" olarak gelişmiştir (Roex ve Degryse, 2004; Çıkrıkçı-Demirtaşlı, 1999).

Bilgisayar Temelli testlerin hızla yaygınlaşması ile birlikte standartlar ve kılavuzlar hazırlanmaya başlanmıştır. Amerika Eğitim Birliği (American Council of Education) 1995 yılında BUT uygulamalarının geliştirilmesi ve eğitimde uygulamalarına ilişkin kılavuz yayınlamış, 2000 yılında Test Yayıncılar Birliği (Association of Test Publisher-ATP) Bilgisayar Temelli Test Kılavuzunu, 2004 yılında Uluslararası Test Komisyonu (International Test Commission-TC) test geliştirenlere, araştırmacılara ve kullanıcılara yönelik olarak “Bilgisayar Temelli ve İnternet Üzerinden Uygulanan Testlere ilişkin Uluslararası Kılavuz” yayınlamıştır. Georgiadou, Triantafillous ve Economides’in (2006) yaptığı çalışmada Rudner’in önerdiği BUT uygulamasını değerlendirme modelini geliştirmiştir (Şekil 5).



Şekil 5. Bilgisayar Uyarlamalı Test Değerlendirme Modeli

Bu modelde BUT uygulamalarını değerlendirmek üzere geliştirilmiş parametreler aşağıda açıklanmaktadır (Georgiadou ve diğerleri, 2006):

Test Parametreleri:

Yararlılık, temel olarak testin amacı ile doğrudan bağlantılıdır ve BUT uygulamalarını değerlendirmede temel bir aşamadır. Bilgisayar sistemlerinin kullanılması her zaman yapılacak sınav için en uygun seçenek olmayabilir (örn: yazılı sınavlar). Sınavı hazırlayanlar değerlendirmenin amacını açık olarak ortaya koyarak BUT uygulaması için yararlılığı tartışmalıdır. Benzer şekilde **geçerlik** ve **güvenirlilik** kanıtları BUT uygulamaları için ortaya konmuş olmalıdır.

Memnuniyet parametresi, bireylerin kullanılan ara yüz ile etkileşimleri ve diğer sistem fonksiyonlarının kullanılabilirliğini içermektedir. **Kullanışlılık**

memnuniyet ile doğrudan bağlantılıdır ve kullanılan terminoloji, sistem tarafından bireye sağlanan geribildirimler, yardımlar, dokümanlar ve ekran tasarımı (renkler, okunabilir font, ikonlar vb.) bu kapsam içinde sayılabilir. BUT uygulamalarında memnuniyet; uygulamanın basit ve açık olması, uygulamanın bireyin sistemin çalışmasına ilişkin zorluk yaşamadan cevaplara odaklanmasını sağlayacak şekilde tasarlanması ile doğrudan ilişkilidir.

Test sonuçlarının bireye nasıl sunulduğuna ilişkin bir parametre olan **raporlama** kapsamında ise BUT uygulamalarının çabuk, doğru, ayrıntılı geribildirimler sağlayacak şekilde bir rapor biçiminde sunabilme durumu değerlendirilmektedir. Bu raporlar: her bir maddeye, her bir içerik alanına yönelik madde gruplarına, aynı güçlük düzeyindeki maddelere, yanlış cevaplanan maddelere ve harcanan zamana ilişkin bilgileri içermelidir.

Test yönetimi, BUT uygulamalarının testi hazırlayan ve uygulayanlar için kullanışlı olması, sistem ve programlama becerilerine sahip olmalarına ihtiyaç duymayacak şekilde hazırlanabilmesi, eğiticilerin testi uygulama, düzenleme ve kullanma konusunda bağımsızlığını arttırmaya yönelik önlemler alınmasını kapsamaktadır.

BUT uygulamaları için, diğer sınavlarda olduğu gibi **test güvenliğini** arttırmaya yönelik önlemlerin, yazılım da dahil olmak üzere uygulamanın her aşaması için tasarlanmış olması gerekmektedir.

Uyumluluk ile; uygulamaya katılan her bir birey için; amaç, bilgi ve tercihleri dikkate alan bir model oluşturulmasına imkan tanıyan uyarlamalı sistemler tanımlanmaktadır. “Herkes için tek tip” uygulamalar olması yerine, tüm öğrenenlerin ihtiyaçlarına ve yeteneklerine uygun esnek bir öğrenme ve değerlendirme ortamının sağlanmasını içermektedir. Ayrıca, web tabanlı uygulamalar uyumlulukla ilgili olarak, kullanıcının ortamını ve kullandığı araçları da (hardware, software, network –ağ bağlantıları) dikkate alınmalıdır.

Soru bankası: BUT uygulamalarının en önemli parçasını, farklı yeterlik düzeylerindeki öğrencilerin değerlendirilmesine olanak sağlayacak farklı madde setlerini içeren soru bankası oluşturmaktadır. Soru bankası, her bir

madde için maddeyi oluşturan metin, doğru cevaba ilişkin detaylı açıklamalar ve madde parametrelerini içermelidir. Ayrıca soru bankası, maddenin ait olduğu içerik alanı, eğitim düzeyi, madde yazarına ilişkin bilgiler, maddenin kullanılma durumu ve kalibrasyonuna ilişkin bilgileri de kapsamalıdır.

Soru bankası, BUT uygulamalarının en önemli bileşeni olmakla birlikte BUT uygulamaları için soru bankasının özelliklerini tanımlama ve geliştirmeye ilişkin süreçler yeterince tanımlanmamıştır. Soru bankasının niteliği temel olarak iki kriter dikkate alınarak değerlendirilebilir (Georgiadou ve diğerleri, 2006):

- Soru bankası, test uygulaması gerekliliklerini yerine getirecek sayıda maddeye sahip olmalı,
- Bankadaki maddeler yeteneğin her düzeyi için yeterli bilgiyi sağlayacak özelliklere sahip olmalıdır. Soru bankasında yeteneğin farklı düzeyleri için bilgi sağlayabilecek farklı güçlük düzeyinde yeterli sayıda madde bulunmalıdır ayrıca madde sayısı maddelerin tekrar sorulma sıklığını belirlediğinden, soru bankası tekrarlanan soruları hatırlamaya izin vermeyecek kadar çok maddeye sahip olmalıdır.

Alan yazında soru bankasının büyüklüğü değerlendirmenin amacı ve oluşturulan testin özelliklerine bağlı olarak değişmekle birlikte nitelikli bir BUT uygulaması için yüksek nitelikte ve iyi bir dağılıma sahip en az 100 madde olması gerektiği belirtilmektedir. Bununla birlikte testi hazırlayanlar, aynı maddeleri sık kullanmak istemiyorlarsa, içerik dengesini en üst düzeyde olmasını istiyorlarsa ve belgelendirmeye yönelik değerlendirme yapıyorlarsa daha çok madde içeren soru bankası gerekli görülmektedir (Georgiadou ve diğerleri, 2006).

Soru bankasında bulunan maddelerin farklı güçlük düzeylerinde olmaları BUT uygulamaları ile farklı yetenek düzeylerinin tümü için doğru kestirimler yapma olanağı sağlar. Madde güçlük düzeyleri farklı yetenek düzeylerine karşılık gelecek yeterli dağılımı göstermediğinde, sabit uzunluktaki BUT uygulamalarında yetenek kestirimlerinde standart hata yüksek olacaktır ya da yetenek kestirimlerini istenen doğrulukta yapabilmek için daha uzun testler oluşturmak gerekecektir. Bu durumda soru bankasının

zayıf kaldığı güçlük ve yetenek düzeyleri için ek soru bankası oluşturmak gerekmektedir. Soru bankası birden fazla içerik alanına ayrılmışsa, bankanın yeterliliği daha zorlayıcı bir durum olarak ortaya çıkmaktadır. Bu durumda her içerik alanı soru bankasındaki madde güçlüğü dağılımını temsil edecek şekilde madde içermelidir.

Bankadaki maddelerin kullanım sıklığı, madde seçim yöntemi ve BUT algoritması ile ilişkilidir. Madde kullanımı, maddelerin çok kullanımı ya da çok az kullanımı gibi iki sorunu içerebilir, Maddelerin çok sık kullanılması hatırlama ya da testi alanların o maddeye ilişkin alana daha çok çalışması ile sonuçlanarak, madde güçlüğü'nün yapay olarak düşmesine neden olabilir. Maddenin çok az kullanılıyor olması da soru bankasının geliştirilmesi için harcanan zaman ve işgücünün iyi kullanılmaması anlamını taşımaktadır.

Bilgisayar Uyarlamalı Test Uygulamaları

BUT uygulamasında; bilgisayar, test maddeleri hakkındaki bilgilerin depolanmasını, üretilmesini, testin uygulanmasını ve puanlanmasını sağlar. Bilgisayar, maddeleri bireye uyarlama işini ardışık olarak yapar (Rudner, 1998; Öztuna, 2008);

- Test, bireyin θ düzeyine ilişkin başlangıç kestirimi ile başlar. Bu tahmin, tüm bireylere sorulan genel bir maddeye verilen yanıt ya da birey hakkında önceden elde edilen bilgiye dayalı olabilir. Bir diğer yolda orta zorluk derecesinde madde seçme yaklaşımıdır.

BUT yönteminde bireyin ilk maddeye verdiği yanıt, sonraki maddenin sıralamasının belirlenmesinde önemli rol oynar. Eğer ilk madde doğru yanıtlandıysa, sonraki madde daha zor, yanlış yanıtlandıysa sonraki madde daha kolay olacaktır. Maddeler bireyin θ düzeyi hakkında en fazla bilgiyi sağlayacak şekilde seçilmektedir.

- İlk maddenin ardından θ düzeyi kestiriminde hangi maddenin en iyi madde olacağı değerlendirilir. Maksimum bilgi yaklaşımı altında, sıradaki en iyi madde, θ düzeyinde en fazla bilgiyi sağlayan maddedir. İlk maddenin

uygulanmasının ardından θ kestirimi için kullanılabilecek kestirim yöntemleri: En Yüksek Olabilirlik (Maximum Likelihood – MLE) ve Bayesian Expected A Posteriori (EAP), Owens's Bayesian Estimation (OWEN) ve Maximum a Posterior Estimation (MAP) yöntemleridir.

- Birey için ilk yetenek kestirimi elde edilince bundan sonra sorulacak maddeleri seçmek için farklı yöntemler kullanılabilir. Bu ikinci aşamada en sık kullanılan madde seçme yöntemi En Yüksek Bilgi (Maximum Information) yöntemidir. Bu yöntemde belirli bir yetenek düzeyi için en yüksek bilgiyi (yetenek düzeyi kestirimi standart hatasında en büyük düşüşü) sağlayacak olan madde seçilir. Bu yöntem dışında Bayesian madde seçim gibi yöntemler ya da yalnızca soruların güçlük düzeyine dayalı seçme yaklaşımları mevcuttur.

- İkinci aşamada seçilen optimal soruya, üçüncü aşamada verilen yanıt, dördüncü aşamadaki θ düzeyini ve güven aralığını yeniden tahmin etmek için kullanılır. Beşinci aşamada, önceden belirlenmiş durdurma kuralının sağlanıp sağlanmadığı kontrol edilir.

- Testi durdurmak için daha önceden belirlenmiş kuralın gereklerinin yerine getirilip getirilmediği kontrol edilir. Durdurma kuralı olarak; zaman, madde sayısı sınırı, θ düzeyine ilişkin tahmin değerindeki değişim, standart hata eşik değeri gibi bir kesinlik göstergesi ya da bu faktörlerin birkaçı bir arada kullanılabilir. Kural bir kere yerine getirildiğinde test sonlandırılır. Durdurma kuralı için belirlenen kural yerine getirilemezse süreç ikinci aşamadan tekrar başlar ve dördüncü aşamada elde edilen θ düzeyi için en çok bilgi sağlayan madde ile devam eder (Linacre, 2000; Öztuna, 2008).

Linacre (2000) BUT uygulamalarında bir testi sonlandırma kararının verilmesinde aşağıda yer alan kuralların da kullanılabileceğini belirtmektedir.

BUT uygulamasına:

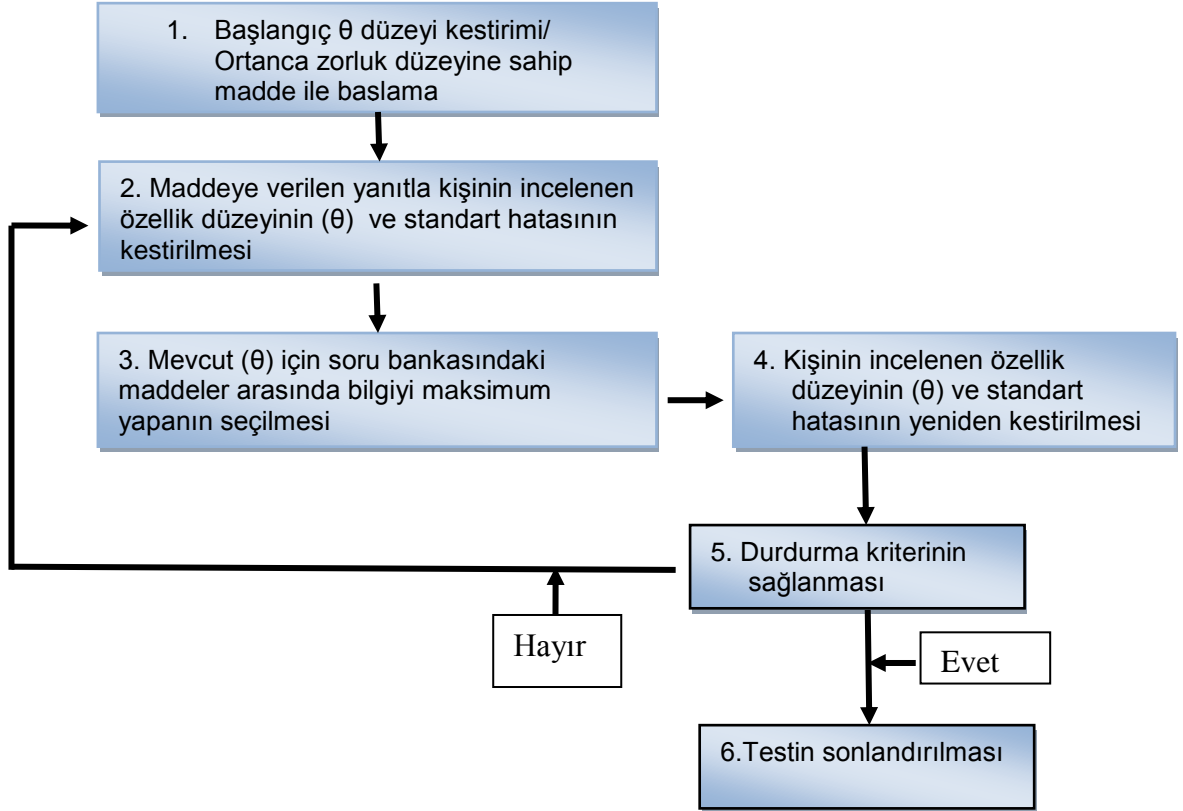
- Soru bankasındaki maddelerin hepsi kullanıldığında
- Maksimum test uzunluğuna ulaşıldığında

- Testi alan birey teste ilgisini kaybettiğini gösteren davranışlar sergilemeye başladığında, örneğin maddelere çok hızlı ya da çok yavaş cevap vermeye başladığında son verilmelidir.

BUT uygulamalarına:

- Birey minimum sayıda test maddesi almadan
- Kapsam içinde yer alan tüm başlıklara ulaşılmadan
- Testin kapsam geçerliliğini sağlayacak yeterlikte madde sorulmadan son verilmemelidir.

BUT yönteminin süreci Şekil 6'da gösterilmiştir (Wainer ve diğerleri, 2000; Bjorner, Chang, Thissen ve Reeve, 2007).



Şekil 6. Bilgisayar Uyarlamalı Test (BUT) Süreci

BUT ilk olarak öğrenci başarısı alanında uygulanmıştır. Son yıllarda ölçmenin diğer alanlarında ve tıp uygulamalarının hastalar üzerindeki sonuçlarına ilişkin tıp araştırmalarında da kullanılmaya başlamıştır (Weiss,

2011). BUT uygulamaları, ölçme değerlendirmeye ilişkin uygulamaların olumlu yönde gelişmesine büyük katkı sağlamaktadır (Kingsbury ve Hourser, 1993; Brown, 1997).

Yurtdışında bilgisayar ortamında bireyselleştirilmiş test uygulamaları konusunda çok geniş bir literatür bulunmaktadır. BUT yöntemi, eğitim alanında kullanılan ve geçerlikleri kanıtlanmış olan testler arasında bulunmaktadır. Bu uygulamalar arasında (Weiss 2003): Eğitimsel Test Servisi (Educational Testing Service- ETS) tarafından yapılan ve Amerika'da lisansüstü eğitim almak isteyen öğrenciler için düzenlenmiş Lisansüstü Giriş Sınavı (Graduate Record Examination- GRE, 1993), yine aynı kuruluş tarafından yapılan ve Lisansüstü Yönetime Geçiş Sınavına Başvuru Kurulu (Graduate Management Admission Council (GMAC) tarafından yönlendirilen İşletme Yüksek Lisansına Kabul Sınavı (Graduate Management Admission Test- GMAT, 1997), Amerika Kolej Sınavı (American College Test- ACT), Eğitim Yetenek Testi (Scholastic Achievement Test- SAT), ana dili İngilizce olmayanlara yönelik olarak İngilizce yeterliğini ölçmek amacıyla yapılan Yabancı Dil olarak İngilizce Sınavı (Test Of English as a Foreign Language- TOEFL, 2000), Silahlı Kuvvetler Mesleki Yetenek Bataryası (Armed Services Vocational Aptitude Battery- ASVAB), ilköğretimde, fen, matematik, dil kullanımı, okuma vb; alanlarda akademik gelişimi değerlendiren Akademik Gelişim Sınavları (Measures of Academic Progress-MAP), yine ilköğretimde matematik yeterliklerini ölçmek amacı ile yapılan Öğrenme ve Öğretme için Matematik Değerlendirme (Mathematics Assessment for Learning and Teaching- MALT) sayılabilir.

BUT uygulamaları ile bu sınavların (GMAT, GRE, TOEFL vb.) daha az soru ile ölçtükleri yetenek düzeyini yüksek güvenirlikle kestirebildiğini destekleyen pek çok kanıt bulunmaktadır. Bu kanıtlara dayalı olarak büyük ölçekte uygulanan pek çok sınavda BUT yöntemi kullanılmaktadır. Örneğin; GRE-BUT uygulamaları 1988 yılında uygulanmaya başlamış, bu aşamada sınav, BUT beraberinde kağıt kalem testi olarak uygulanmaya devam etmiştir. Bu süreçte BUT uygulamalarından elde edilen puanlar, kağıt kalem testinden elde edilen veriler ile karşılaştırılmış, kağıt kalem testi uygulamaları

için gerekenin yarısı kadar soru ile değerlendirme yapılabildiği ortaya konmuştur. Bu çalışmalara dayalı olarak 1993 yılında GREBUT uygulamaları başlatılmıştır (van der Linden ve Glas, 2000). Benzer şekilde BUT uygulamaları ile GMAT ve TOEFL'un da daha az soru ile ölçtükleri yetenek düzeyini yüksek güvenirlikle kestirebildiği gösterilmektedir (GRE, 2013; GMAT, 2013, TOEFL, 2013).

Wang ve Kolen (2001) yılında yapmış oldukları çalışmalarında BUT ve kağıt-kalem testi arasındaki karşılaştırmaların genel olarak, geçerlik, psikometrik özellikler ve testin uygulanmasına yönelik kriterleri içerdiğini belirtmektedirler. Weiss (2010) alan yazında BUT'a ilişkin yapılmış çalışmaları şu başlıklar altında toplamıştır: İçerik/kapsam dengesi, kestirim yöntemleri, durdurma kuralları, çoklu ölçekleme, test soruları ve simülasyon çalışmaları.

BUT uygulamaları, kağıt kalem testi uygulamalarıyla elde edilen kestirimleri çok daha az sayıda soru ile elde edebilmektedir. Kağıt kalem testi olarak uygulanan kısa testlerin aksine, BUT ölçmede doğruluk ve güvenirlikten pek fazla taviz vermeden, daha az soru ile bireyin yetenek düzeyini kestirerek, cevaplayıcı ve uygulayıcıların iş yüklerini azaltmaktadır (Zitny, Halama, Jelinek ve Kveton, 2012).

BUT uygulamalarının ilk yıllarında soru sayısı ve yetenek kestirimlerine ilişkin yapılan çalışmalar, BUT uygulamaları ile geleneksel test uygulamaları ile elde edilen ölçme doğruluğuna %50 daha az soru ile ulaşıldığını göstermiş, bu sonuçlar da BUT uygulamalarının kullanımını ve yaygınlığının artmasını desteklemiştir.

Yakın dönemlerde yapılan çalışmalarda, tipik performansın ölçülmesi ve sağlık alanında, BUT uygulamaları ile ölçmenin doğruluğundan taviz vermeden kağıt kalem uygulamalarındaki test uzunluğunda %95 azalma sağlandığı gösterilmektedir (Weiss, 2011),

Kaptan (1993) yapmış olduğu çalışmasında kağıt kalem testleriyle bilgisayar ortamında bireyselleştirilmiş testleri karşılaştırmıştır. Bu çalışmada

ÖSS sorularından elde edilmiş matematik soruları kullanılmıştır. Öğrencilere kağıt kalem testinde 50 soru, bilgisayar uyarlamalı testte ise 14 soru uygulanmıştır. Çalışma sonucunda test uygulama süresi ve uygulanan soru sayısında %70'lik düşüş saptanmış, yetenek kestirimleri arasında istatistiksel olarak manidar bir fark bulunmamıştır.

İşeri (2002) Ortaöğretim Kurumları Seçme Yerleştirme Sınavı ve Özel Okullar Sınavının matematik bölümünü kullanarak soru bankası oluşturmuştur. Araştırmacı bu soru bankasını kullanarak öğrencilerin matematik başarılarını bilgisayar ortamında bireyselleştirilmiş test uygulaması ile ölçmüş ve bilgisayar ortamında bireyselleştirilmiş testlerin öğrencilerin yetenek düzeyini daha az soru kullanarak güvenilir biçimde ölçtüğü sonucuna varmıştır.

Kalender (2009) ÖSS fen alt testi için CAT uygulanabilirliğini araştırmıştır. Çalışmasında, 2005 ve 2006 yılları için Anadolu Lisesi, devlet ve özel okul türleri ile ayrı ayrı, 3000 gerçek öğrenci verisi üzerinde post-hoc simülasyonu yöntemi ile CAT uygulaması yapmış ve bu yöntemin öğrencilerin ÖSS fen alt testindeki yetenek düzeylerini kestirme gücünü ortaya koymaya çalışmıştır. Elde edilen sonuçlara göre en ideal CAT formatının tüm okul türleri için EAP yöntemi ile 0,3 standart hata eşik değerine sahip olan bir uygulama olduğu belirlenmiştir. EAP/0,3 SE yöntemi kestirimleri ile gerçek ÖSS kestirimleri arasındaki korelasyonların ortalaması 0,95 ve gerekli olan soru sayısı ortalamaları 2005 yılı için 17,3; 2006 yılı için ise 19,5 olarak hesaplanmıştır. Çalışmada, ÖSS fen alt testinde öğrencilerin CAT formatı ile elde edilen yetenek kestirimleri ile gerçek ÖSS'deki yetenek kestirimleri arasındaki ilişkiler incelenmiştir. Bu amaçla, ODTÜ Hazırlık Okulu'nda okuyan 33 öğrenci, geçmiş yıllara ait soruları içeren CAT uygulaması formatında ÖSS fen alt testi almışlar ve gerçek ÖSS'deki yetenek kestirimleri ile CAT uygulaması kestirimi arasındaki korelasyon katsayısı 0,74 olarak bulunmuştur.

Miller (2003) yürütmüş olduğu çalışmasında, eyalet düzeyinde öğrenci değerlendirmelerinde BUT uygulamalarının yeterlik kestirmedeki güvenilirliği ve öğrencilerin BUT ve kağıt kalem testi uygulamalarına ilişkin tutumlarını

değerlendirmiştir. Kağıt kalem testi olarak uygulanan Michigan Eğitim Değerlendirme Programı (Michigan Education Assessment Program-MEAP) okuma testi puanları, BUT uygulama sonuçları ve Ulusal Eğitim Testleri ve Kamu Politikaları Kurumu (National Board on Educational Testing and Public Policy) tarafından geliştirilen bir ölçekten elde edilen veriler, çalışmanın verilerini oluşturmaktadır. Çalışmaya, Michigan Eğitim Değerlendirme Programı iki okuma alt testini, BUT uygulamasını ve tutum ölçeğini tamamlayan kentsel bölge okullarının dördüncü sınıfında okuyan 267 öğrenci dahil edilmiştir. Araştırmada BUT alt testlerinin sonuçları ile kağıt kalem testi sonuçları arasındaki ilişkinin sayısal değeri istatistiksel olarak manidar bulunmuştur.

Zitny (2011) yaptığı çalışmada başarı testleri, yetenek testleri, klinik psikoloji testleri, kişilik testleri ve sağlık alanında yapılmış BUT uygulamalarına ilişkin 15 çalışmayı güvenirlik, kullanılabilirlik (soru sayısının azalması kapsamında) ve geçerlik (var olan araçlarla korelasyon) bağlamında değerlendirmişler ve BUT uygulamalarının zaman ve soru sayısını azaltarak optimal düzeyde bilgi sağladığı sonucuna ulaşmışlardır. Araştırmacılar, BUT uygulamaları ve soruların tamamının uygulanması ile elde edilen puanlar arasında yüksek korelasyon ($r=0,83-0,99$) bulunduğunu ve BUT'un geçerlik, güvenirlik düzeylerinin hali hazırda kullanılan araçların geçerlik ve güvenirlik kanıtları ile uyumlu olduğunu ($r=0,58-0,83$) göstermişlerdir.

Zitny ve diğerleri (2012), bilişsel yeterlik testlerinin kağıt kalem, bilgisayar temelli ve BUT yöntemleri ile uygulamalarından elde edilen geçerliklerini ve farklı yöntemlerle yapılan sınavların uygulanması için gerekli soru sayısını karşılaştırmak amacıyla bir çalışma yürütmüşlerdir. Bu çalışmada 803 ortaokul öğrencisinin verilerini kullanmışlardır. Çalışmada, kağıt kalem testi olarak uygulanan Algılama Potansiyeli Testi (Test of Intellect Potential-TIP) ve Viyana Matrisler Testi (Vienna Matrices Test-VMT) setleri, bilgisayar temelli olarak gerçek uygulamalardan elde veriler ve simüle BUT uygulama verileri olmak üzere üç veri seti kullanılmıştır. Çalışmada, 567 öğrenci kağıt kalem testi, 236 öğrenci ise Bilgisayar temelli ve BUT

uygulamaları ile TIP ve VMT sorularını cevaplamışlardır. Sonuçlar, BUT uygulamaları ile elde edilen geçerlik düzeylerinin (kriter ve yapı geçerliği) diğer yöntemler ile elde edilenlerle kabul edilebilir düzeyde uyuşmakta olduğunu göstermektedir. BUT uygulamalarının daha az sayıda soru ile benzer geçerlik düzeylerini sağlamakta olduğu rapor edilmiştir. BUT uygulamalarında geleneksel yöntemlere göre TIP’de %55, VMT’de %54 oranında daha az soru kullanılmıştır.

Bulut ve Kan (2012), BUT yönteminin Akademik Personel ve Lisansüstü Eğitimi Giriş Sınavına (ALES) uygunluğunu araştırdıkları çalışmalarında, hali hazırda kullanılmakta olan klasik formları ve BUT uygulamalarından elde edilen yetenek kestirimlerini karşılaştırmak amacı ile post-hoc simülasyonlar uygulamışlardır. Veri setini, sınava tüm katılanlar arasından rastgele örnekleme seçilmiş olan 10,000 kişiye 2008 yılında uygulanmış olan ALES verileri oluşturmaktadır. Araştırmacılar, EAP yetenek kestirimi yöntemi ile 0.25, 0.30 ve 0.40 standart hata eşik değeri ile BUT uygulanabileceğini, BUT ile klasik formlardan elde edilen yetenek kestirimleri arasındaki korelasyonunun 0.93 ve üzeri olduğunu bulmuşlardır. BUT uygulamasında kullanılan soru sayısı her bir alt test için 9 ila 22 arasında değişmektedir. BUT uygulamaları ile ALES’te tüm soruların uygulandığı durumdaki kadar net yetenek kestirimi sağlanırken soru sayısı %70 azalmıştır.

Eggen ve Straetmans (2000) Hollanda’da öğrencileri BUT uygulamaları ile yetenek düzeylerine göre farklı başarı grubuna ayırmak üzere yapmış oldukları çalışmalarında, kağıt kalem uygulamalarına göre BUT uygulamalarında soru sayısında %22-44 oranında azalma olduğunu saptamışlardır.

Tıp Eğitiminde Bilgisayar Uyarlamalı Test

Tıp eğitiminde bilgisayarın pek çok eğitim ortamında farklı amaçlarla kullanımı yaygındır. Son yıllarda kağıt kalem kullanımına dayanan pek çok işlem örneğin kliniklerde hasta kayıtlarının elektronik olarak tutulması, kitapların elektronik ortamlara aktarılması vb. yerini bilgisayara bırakmıştır.

Bu yönede hızlı değışime ve bilgisayar ortamında değerdirmelerin avantajlarının ortaya konmuş olmasına rağmen pek çok tıp faköltesinde hala yaygın olarak kağıt kalem testleri kullanılmaktadır (Hols-Elders ve diğderleri, 2008).

Benzer şekilde Tıp Eğıtinde de bilgisayar ortamında değerdirmenin avantajları tanımlanmıştır; Bilgisayar ortamında testler zaman ve yerden bağımsız olarak yürütölebilmektedir. Sonuęlar değerdirmenin hemen sonunda öğrencilere duyurulabilmektedir. Soruların parametrik özelliklerinin hesaplandığı bir soru bankasının kullanılması, ölçmede geçerlik ve güvenirlik düzeyini arttırmaktadır. Ayrıca bilgisayar ortamında yapılan testler, soruların amacına uygun olarak birbirini izlemesi, hareketsiz ya da hareketli görseller, ve kalp sesi gibi sesli aktarımlar kullanabilme, testi öğrencinin daha önceki performansına göre yapılandırma ve her bir öğrenciye özgü geribildirim sağlayabilme gibi pek çok avantaja sahiptir. Yine tıp eğitiminde bilgisayar ortamında iletişim becerileri gibi beceri değerdirmeleri de yapılabilmektedir (Hols-Elders ve diğderleri, 2008).

Öğrencilerin bilgisayar ortamında yapılan ölçmelerden memnuniyetlerini belirlemeye yönelik olarak yapılan çalışmalarda menuniyetin yüksek oluşu belirtilmektedir. Bilgisayar kullanımı, puanların ve soru ve test istatistiklerinin otomatik olarak hesaplanabilmesine olanak sağlayarak, testin niteliğini arttırmakta, testin yürütölmesine ilişkin eğitici yükünü hafifletmekte ve zaman kaybını azaltmaktadır (Hols-Elders ve diğderleri, 2008).

Karay ve Schaubert, Stoch ve Schuettpeiz-Brauns (2012) tarafından Cologna Üniversitesi Tıp Faköltesinde (Almanya) yapılan çalışmada Gelişim Sınavı kağıt kalem testi formatında ve bilgisayar temelli uygulanmış ve testin öğrenciler tarafından kabulunu değerdirmek üzere bu iki uygulamayı karşılaştırmışlardır. Cologna Üniversitesi Tıp Faköltesinde gelişim sınavı 2003 yılından bu yana biçimlendirmeye yönelik değerdirmeye amacı ile yılda iki kez uygulanmakta ve her test tek doğru cevaplı 200 çoktan seçmeli soru içermektedir. Test süresi üç saattir. Çalışmada, kağıt kalem formatında iki test (n=1264, 2008 güz dönemi; n=1333, 2009 bahar dönemi) ve Bilgisayar ortamından iki test (n=1416, 2009 güz dönemi; n=1296, 2010 bahar dönemi)

olmak üzere toplam dört uygulama ve 5309 katılımcının sonuçları değerlendirilmiştir. Her değerlendirmenin sonunda gönüllük esasına dayalı olarak öğrenciler, sınavın kabul edilme düzeyine ilişkin beşli dereceleme ölçeği ile (1 çok iyi, 5 çok kötü) cevaplanan sorular içeren bir anketi çevrimiçi olarak doldurmuşlardır. Ayrıca aynı uygulamada açık uçlu sorular ile öğrencilerin görüş ve önerileri alınmıştır. Beşli dereceleme ölçeği kullanılarak puanlanan sorularda ortalama, açık uçlu sorularda ise negatif ve pozitif görüşlerin sayısı değerlendirilmiştir. Sonuçlar, klinik öncesi dönemde değerlendirmenin farklı formatları arasında kabul görme açısından bir fark bulunmadığını, klinik dönemde ise Bilgisayar Temelli değerlendirmenin daha fazla kabul gördüğünü göstermektedir. Bilgisayar Temelli değerlendirme ile anında geribildirim verilmesinin bu kabulü arttırdığı ve bu yolla da biçimlendirmeye yönelik değerlendirmenin yararlılığının arttığı sonucuna varılmıştır.

Roex ve Degryse'nin (2004), genişletilmiş eşleştirmeli soruların BUT uygulamalarında kullanılması ve kapsam geçerliliğini araştırdıkları çalışmalarında, Belçika/Flanderste aile hekimliği sertifikasyonu için yapılan ve modüler bir yapı içeren sınav uygulamasında BUT yöntemi kullanmış ve etkililiği incelenmiştir. Eğitimlerinin sonunda öğrencilere sözlü sınav, Nesnel Yapılandırılmış Klinik Sınav ve yazılı sınav uygulanmıştır. Yazılı sınavlar 165 genişletilmiş eşleştirmeli sorudan oluşmaktadır. Sorular genel tıp kapsamını içerecek şekilde hazırlanmış bir belirtke tablosu kullanılarak belirlenmektedir. Sınav süresi üç saattir ve öğrencilerin tümü aynı soruları cevaplamaktadır. Daha önce sorulmuş soruları içeren (son altı yıl) geniş bir soru bankası bulunmaktadır. Araştırmacılar aile hekimliği kapsamında yürütülen bu sınavların BUT uygulamaları açısından tartışmalı iki alanı olduğunu belirtmektedirler. Bunlardan birincisi, BUT uygulamaları için soru bankasının tek boyutlu olması gerekliliği göz önüne alındığında aile hekimliği alanının farklı alt alanlara sahip olması, diğeri ise olgu özerkliğinin (case specificity) olmasıdır. Olgu özerkliği, bir olgu için elde edilmiş puanların, diğeri bir olguyu içeren sorular için bir puan kestirimi sağlayamamasıdır. Farklı alt alanların olduğu bu gibi durumlarda kapsam geçerliliğini ve test güvenirliğini arttırmaya yönelik olarak içerik/kapsam alanları için geniş bir örneklem

seçilmesi ve alt alan için özel testler geliştirilmesi önerilmektedir (Van der Vleuten ve diğerleri, 1996a). Bu durum mezuniyet sonrası tıp eğitimi kapsamında yapılan değerlendirmelerde BUT uygulamaları geniş bir soru bankasını gerektirmekte ve testin uzunluğunu arttırmaktadır. Araştırmacılar 47 öğrenci ile, farklı uzunlukta (20- 50 soru) BUT uygulaması gerçekleştirmişler ve testler için farklı süreler belirlemişlerdir. Soru sayısı, test süresi ve yetenek kestirimlerini incelediklerinde, soru sayısı ve bireylerin yetenek düzeyleri arasında pozitif bir ilişki bulmuşlardır. Testin kalitesini ve ölçmenin öğrenmeye etkisini arttırmak amacıyla kapsamlı, alt alanlara göre dengeli bir soru bankasının gerekli olduğunu belirtmişlerdir.

Kreiter, Ferguson ve Gruppen (1999)'da yaptıkları çalışmada üçüncü sınıf iç hastalıkları dersi kapsamında üçüncü sınıf öğrencileri ile bilgisayar ortamında uygulanan sınavı, bireye uyarlanmış teste dönüştürerek etkinliğini değerlendirmeyi amaçlamışlardır. Çalışmada 200 genişletilmiş eşleştirmeli soru kullanılmıştır. Çalışmaya toplam 631 öğrenci katılmış olup öğrencilere önce 200 soruluk kağıt kalem testi uygulanmıştır. Soruların parametre kestirimleri (ortalama madde gücü= 0.88) ve testin güvenirliği hesaplanmıştır (Cronbach alfa=0.88). Daha sonra gerçek uygulama verileri kullanılarak yapılan simüle BUT uygulamasından veriler elde edilmiştir. Bu uygulamada ortalama madde gücü parametresi (ortalama -1.47), madde ayırt edicilik parametresi (ortalama 0.557) ve farklı güvenirlik düzeyleri için, gerekli olan soru sayıları hesaplanmıştır (0.75 güvenirlik düzeyi için, yaklaşık 52 madde). Bu iki farklı uygulamanın sonuçları karşılaştırılmış ve sonuç olarak, 200 soruluk kağıt kalem testinden elde edilen ölçmenin güvenirliğine BUT uygulamalarında, soruların yaklaşık yarısı ile ulaşabildiği gösterilmiştir. Tıp eğitiminde, kağıt kalem testleri için kullanılan sorulardan, BUT uygulamaları için soru bankaları hazırlanarak kullanılması avantajlı bulunmuştur.

Araştırmalar BUT uygulamaları ile daha az madde ile yetenek düzeylerini yüksek güvenirlikle kestirilebildiğini, kağıt-kalem uygulamaları ile elde edilen ölçme doğruluğuna çok daha az madde ile ulaşılabilirdiğini göstermektedir.

BÖLÜM III

YÖNTEM

Bu bölümde, araştırmanın modeli, araştırma grubu, veriler, veri toplanması ve verilerin analizi yer almaktadır.

Araştırma Modeli

Bu çalışmada, gelişim testi için bir soru bankasının oluşturulması ve bu soru bankası aracılığı ile BUT yönteminin uygulanabilirliği araştırılmıştır. Bu amaç çerçevesinde, farklı grupların karşılaştırılmasına dayanan uygulamalı bir araştırmadır.

Çalışma Grubu

Çalışmanın grubunu Ankara Üniversitesi Tıp Fakültesi 2010-2011 eğitim-öğretim yılı Dönem 1-5 öğrencileri oluşturmaktadır. Gönüllülük esasına dayalı katılımla gerçekleştirilen gelişim sınavında öğrencilerin büyük çoğunluğunun cevaplarına ulaşılmıştır (%89.7).

Veri Toplama Aracı ve Verilerin Toplanması

Araştırma verileri, Ankara Üniversitesi Tıp Fakültesinde 2010 - 2011 eğitim-öğretim yılında gerçekleştirilmiş olan Gelişim Sınavı maddelerine öğrencilerin vermiş oldukları cevaplardan oluşmaktadır.

Ankara Üniversitesi Tıp Fakültesi'nde Gelişim Sınavı Uygulaması

Ankara Üniversitesi Tıp Fakültesinde, 24/03/2007 tarihinde ilk kez Gelişim Sınavı uygulanmıştır. İkinci GS, 2007-2008 eğitim öğretim yılında, sadece Dönem VI'ların katılımı ile gerçekleştirilmiştir (Çizelge 1). 2008-2009 eğitim öğretim yılında ise yönetmelik değişikliği ile gelişim sınavındaki başarının öğrencinin başarı notuna katkısı sağlanmış, Gelişim Sınavına bu öğretim yılında, büyük kısmını ilk üç dönem öğrencilerinin oluşturduğu toplam 800 öğrenci katılmıştır. 2009-2010 eğitim öğretim yılında yılında yapılmış olan gelişim sınavına 1089 öğrenci, 2010-2011 eğitim öğretim yılında ise 1206 öğrenci katılmıştır.

Çizelge 1. Ankara Üniversitesi Tıp Fakültesi Gelişim Sınavına Katılan Öğrenci Sayısının Yıllara Göre Dağılımı ve Toplam Öğrenci Sayısı İçindeki Yüzdesi

Eğitim Öğretim Yılı	Katılan Öğrenci	
	Sınava Katılan Öğrenci Sayısı (n)	Sınava Katılma Yüzdesi %
2006-2007	520	29,3
2007-2008	98*	-
2008-2009	800	46
2009-2010	1089	58,9
2010-2011	1206	86,7

* Sadece Dönem IV öğrencileri sınava dahil edilmiştir.

Ankara Üniversitesi Tıp Fakültesi'nde gelişim sınavı uygulama aşamaları;

i) Sınav yapılandırma ve hazırlık dönemi

Ankara Üniversitesi Tıp Fakültesinde gelişim sınavlarında çoktan seçmeli 200 madde bulunmaktadır. Sınav içeriği tıp öğrencilerinin mezun olduklarında edinmeleri gereken bilgi bütünlüğü göz önünde tutularak

belirlenmiştir. Sınav maddelerinin anabilim dalları ve organ-sistemlere göre dağılımını ve ağırlığını düzenlemek için çok yönlü bir matris oluşturulmuştur.

Bu matrisin oluşturulmasında Türkiye’de sık görülen hastalıklar ve demografik özellikler göz önünde bulundurulmuştur. Ayrıca, madde seçimi için; yurt dışından örnekler Amerika Birleşik Devletleri Tıp Lisans Sınavı (United States Medical Licensing Examination-USMLE), Maastricht Üniversitesi gelişim sınavı madde dağılımı, Ulusal Çekirdek Eğitim Programı (ÇEP), Uluslararası Hastalık Sınıflaması-10 (International Classification of Disease-10-ICD-10), Sağlık Bakanlığı sık görülen hastalıklar listesi, Tıpta Uzmanlık Sınavı (TUS) madde dağılımı dikkate alınmaktadır.

ii) Maddelerin hazırlanması ve toplanması:

Öğretim üyelerine gelişim sınavı hakkında bilgilendirici bir metin, test maddesi hazırlama ile ilgili temel bilgilere ilişkin bir bilgilendirme yazısı ve madde örnekleri iletilmekte ve her Anabilim dalından belirlenen sayı ve özelliklerde gelişim sınavı maddesi istenmektedir. Hazırlanan sorular Anabilim Dalları Akademik kurulların değerlendirilmektedir. ve Gelen sorular Tıp Eğitimi Anabilim Dalı tarafından değerlendirilerek Anabilim/bilim dalı ağırlıkları ve sistemlere göre dağılım göz önünde bulundurularak sınav için madde seçilmektedir.

iii) Sınavın uygulanması:

- Salon bulabilmek için sınav tarihi belirlenirken ÖSYM sınav takvimi göz önünde bulundurulmaktadır.
- Tüm öğrencilerin aynı anda sınava girecekleri göz önünde bulundurularak hazırlanan sınav için Ankara Üniversitesi Tıp Fakültesi dışında farklı fakültelerden/Üniversitelerden yardım alınmakta ve pek çok sınav salonu hazırlanmaktadır (örn; 2007-2008 yılında Ankara Üniversitesi Siyasal Bilgiler Fakültesi, Ankara Üniversitesi Hukuk Fakültesi, İletişim Fakültesi ve Ankara Üniversitesi Tıp Fakültesi’nde toplam 39 salon kullanılmıştır).

iv) Değerlendirme:

Gelişim sınavı sonuçları, bağıl ölçüt esas alınarak (Genel ve dönem içinde) değerlendirilmektedir. Öğrencilere, sınav sonuçları, bireysel olarak, internet sayfası ve e-posta ile duyurulmaktadır. Öğrencilere, hem katılan tüm öğrenciler arasındaki hem de kendi dönemi içindeki durumları, sistemler ve anabilim dallarına göre ayrıntıları da içeren özel bir karne ile iletilmektedir. Karne ekinde öğrenciye yanlış yaptığı maddeleri için de geri bildirim verilmekte, ayrıca, her Anabilim dalına genel sınav sonucu ve kendi maddeleri ile ilgili tüm değerlendirmeleri içeren bir rapor gönderilmektedir.

2010-2011 eğitim öğretim yılında Anabilim Dallarından gelen sorular Anabilim/bilim dalı ağırlıkları, sistemlere göre dağılım göz önünde bulundurularak Tıp Eğitimi Anabilim Dalı tarafından değerlendirilmiş ve sınav için 200 soru seçilmiştir. Sınav 14 Mayıs 2011 tarihinde Ankara Üniversitesi Tıp Fakültesinin Morfoloji, Cebeci ve İbn-i Sina Yerleşkelerinde bulunan 30 amfi/derslik/salon kullanılarak gerçekleştirilmiştir. 200 sorudan oluşan sınavın süresi 180 dakika olarak belirlenmiştir. Sınava girmek için başvuru yapan 1264 öğrenci içinden 1206 öğrenci sınava katılmıştır.

Verilerin Analizi

Çalışmanın veri analizi üç aşamada gerçekleştirilmiştir:

1. Soru bankasının oluşturulması; soru bankasında bulunan 103 madde ile geçerli ve güvenilir ölçme yapıp yapılamayacağının sınanması
2. Rasch analizi sonucunda elde edilen madde parametreleri kullanılarak SmartCAT aracılığıyla veri türetilmesi ve bu veriler kullanılarak simüle BUT uygulaması
3. Bireylere ilişkin BUT yöntemi ile elde edilen yetenek düzeyleri (θ_{BUT}) ile tüm maddeler üzerinden Rasch analizinden elde edilen yetenek düzeyleri (θ_{Rasch}) arasındaki uyumun incelenmesi

1. Soru bankasının oluşturulması

BUT uygulanmadan önce soru bankasının psikometrik özelliklerinin değerlendirilmesi gerekmektedir. Soru bankasında bulunması istenen özellikler; soru bankasının tek boyutlu bir yapıya sahip olması, soru bankasındaki maddelerin incelenen Madde Yanıt Fonksiyonu (MYT) modeline uyum göstermesi ve Maddenin Farklı Fonksiyonlaşması (MFF) göstermemesidir.

Bu aşamada maddelerin Rasch modeline uyumu, tek boyutluluk varsayımı ve MFF özelliği incelenerek soru bankasının içsel yapı geçerliği (internal construct validity) sınanmıştır.

1.1 Kayıp veriler için değer atama

Çoktan seçmeli soru formatının yapısı gereği seçenekler içermesi, o maddeyle ölçülen özelliğe sahip olmayan veya kısmen sahip olan yanıtlayıcıyı, şansını kullanarak doğru yanıtı bulmaya yönltebilir. Şans başarısı olarak tanımlanan bu durum testin geçerlik ve güvenilirliğini olumsuz etkilemektedir (Tekin, 2009).

Şans başarısını azaltmak için yanlış cevaplara yönelik olarak düzeltme formülü kullanılması ve farklı dönem öğrencilerinin aynı sınavı almaları çok sayıda kayıp veri olmasına neden olmaktadır. Bu nedenle, verilerin analizine geçmeden önce, kayıp verilerin çok olmasının parametre kestirimleri üzerinde yaratacağı; standart hatanın yüksek çıkması, parametre kestiriminin yapılamaması gibi olumsuzlukları gidermek üzere Sijtsma ve van der Ark, (2003) tarafından önerilen Yanıt Fonksiyonu (YF) ile değer atama yoluna gidilmiştir. Öğrencinin yetenek düzeyini gösteren parametre θ ve madde zorluğuna ait parametre ise β ile gösterildiğinde Rasch modelinde öğrencinin β_j zorluğuna sahip j 'inci maddeden y_j puanını alma olasılığı madde yanıt fonksiyonu (MYF) ile tanımlanır. Sijtsma ve van der Ark (2003) tarafından önerilen Yanıt Fonksiyonu MYF'nin kestirimidir ve bu kestirim kullanılarak kayıp veriler yerine değer ataması yapılır. MYF'nin kestirimini yapabilmek

için, öğrencinin değer ataması yapılacak j'inci madde dışında kalan K-1 (K: toplam madde sayısı) tane madde üzerinden toplam puanı kullanılır. Burada, öğrencinin bilgi düzeyi parametresi θ 'nın tahmini olarak kullanılmaktadır. MYF'nin kestirimi olarak YF kullanılır. Böylelikle, öğrencinin bazı madde ya da maddelere vermediği yanıtlar YF kullanılarak tahmin edilir (Sijtsma ve van der Ark, 2003).

Doğanay Erdoğan B. (2012) çalışmasında, kayıp verilerin ve kayıp veriler yerine çoklu değer atama yöntemlerinin Rasch modellerinden bulunan kişi ve madde parametre kestirimleri üzerindeki etkisini incelemiştir. Çalışmanın sonucunda, yanıt fonksiyonu ile değer atamanın, diğer yöntemler arasında en iyi performansa sahip olduğu, ölçeklerin Rasch modelleri ile değerlendirilmesinde veride kayıp gözlem olduğunda kullanılabileceği gösterilmiştir.

Bu çalışmada, YF değer ataması için, van Ginkel ve van der Ark (2005) tarafından geliştirilen SPSS syntax'ı kullanılmıştır. Mevcut kişi ve madde parametreleri kestirimlerini önemli derecede değiştirmeyecek şekilde değer ataması yapılmıştır.

1.2 Soru bankasının Psikometrik özellikleri (Soru Bankalarının İçsel Yapı Geçerliliği-Internal Construct Validity ve Güvenirlik)

Eğitim bilimlerinde yaygın olarak kullanılan, iki kategorili yanıtlar için geliştirilmiş olan tek parametrelili Rasch modeli, bu çalışma kapsamında elde edilen soru bankasının psikometrik özelliklerinin incelenmesinde kullanılmıştır. Araştırmada elde edilen soru bankasının psikometrik özellikleri RUMM 2020 (RUMM Laboratory Pty. Ltd., 2007) paket programı yardımıyla aşağıdaki protokol uygulanarak incelenmiştir:

- Yanıtlar için seçenek analizinin yapılması
- Tek boyutluluk ve yerel bağımsızlık varsayımlarının test edilmesi
- Ölçme aracındaki maddelerin modele uyumunun incelenmesi

- Maddelerin değişmezlik özelliğini sağlayıp sağlamadığının test edilmesi
- Maddenin Farklı Fonksiyonlaşması (MFF) işlev farklılığının test edilmesi (Pallant ve Tennant, 2007; Öztuna 2008).

Bu çalışmada elde edilen soru bankasında bulunan 103 madden oluşan testin/Soru bankasının iç tutarlığı ve güvenirliği Rasch analizinden elde edilen Birey Ayırsama İndeksi (BAI) [Person Separation Index-(PSI)] ve KR-20 ile incelenmiştir.

1.2.1 Seçenek Analizi

Madde testin puanlanabilen en küçük birimdir. Bir çoktan seçmeli madde, biri kök, diğeri seçenekler olmak üzere iki kısımdan oluşur. *Kök*, sorunun sorulduğu, diğeri bir deyişle davranışın mevcut olması halinde ortaya çıkmasını sağlayacak uyarıcının sunulduğu kısımdır. Maddenin diğeri kısmı seçeneklerdir. *Seçenekler* kökte sorulan soruya cevap olduğu düşünülebilecek olan biri doğru, diğeri doğru olmayan önerilerdir. Doğru olmayan seçeneklere *çeldirici* denir. *Çeldirici* doğru cevabı bilmeyen cevaplayıcılara doğruymuş gibi görünmek üzere yazılmış cevaplardır.

Hazırlanan bir testin uygulanmasından sonra madde güçlüğüne, maddenin ayırt ediciliğine ve seçeneklerin çeldiriciliğine bakılmasına testin “madde analizi” denir.

“0-1” olarak puanlanan çoktan seçmeli maddelerle yapılan bir değerlendirmede bir maddeyi anahtara göre doğru işaretleyen öğrenci o maddeyi doğru cevaplamış kabul edilir ve bireye o madde için “1” puan verilir; maddeyi boş bırakan veya çeldiricilerden birini işaretleyen öğrenci ise o maddeyi yanlış cevaplamış kabul edilir ve bireye o madde için “0” puan verilir. Buradaki temel ilke, sadece tam bilgiye sahip bireylerin doğru cevaplarına ve tahminle verilen doğru cevaplara puan atamaktır. Tamamen yanlış bilgiye sahip olan, kısmi bilgiye sahip olan ve tahminle cevaplama davranışında bulunarak yanlış cevap veren bireylerin cevapları ise yanlış

cevap kategorisinde değerlendirilmektedir (Jaradat ve Tollefson, 1988). Bu iki kategorili puanlama, cevaplayıcının madde içeriğini ya hiç bilmediği ya da tam olarak bu bilgiye sahip olduğu varsayımına dayanır.

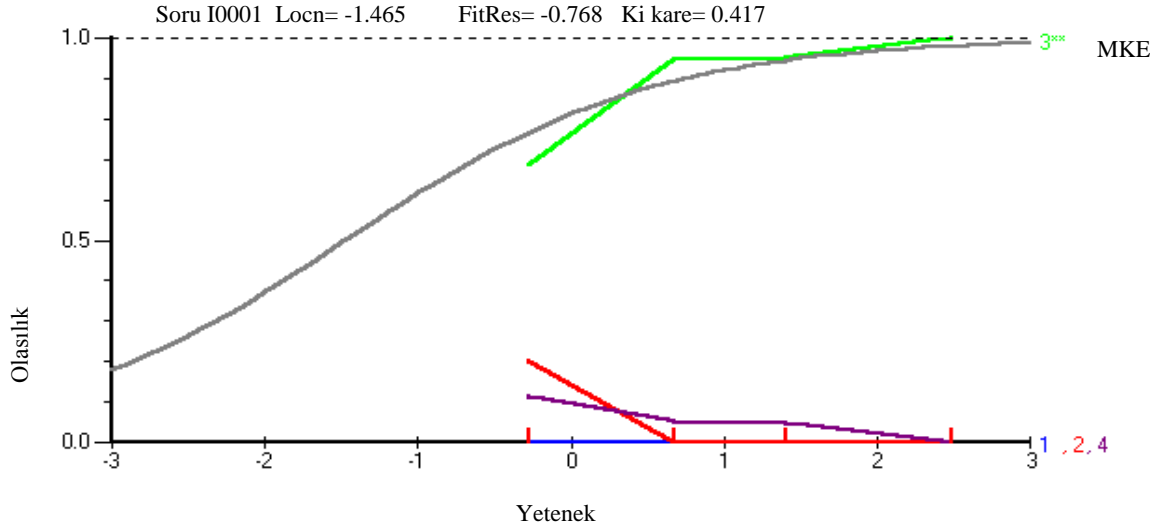
Bir madde hakkında tam bilgi edinmek ve o maddeyi daha iyi hale getirmek için seçenek analizlerinin de değerlendirilmesi gereklidir. Seçenek analizi cevaplayıcıların ölçülmek istenen değişkene ilişkin öğrenmeleri hakkında bilgi sağlar. Bu doğrultuda maddelere verilen yanıtlara bakılarak seçeneklerin işlevliliği incelenir, yanlış cevapların çeldiricilere dağılımı değerlendirilir. Maddenin yanlış cevaplarının, çeldiricilere eşit olarak dağıtılması beklenir (Asril ve Marais, 2011).

Bu çalışma kapsamında yapılan analizde, her madde için madde karakteristik eğrisi ile birlikte her bir çeldiricinin dağılımı elde edilmiş ve çeldirici örüntüleri değerlendirilmiştir. Doğru seçenek, madde karakteristik eğrisini benzer şekilde takip etmelidir. Çeldiricileri oluşturan diğer seçeneklerin gözlenme oranı, düşük yetenek düzeyindeki bireyler için yüksek olması beklenirken, yetenek düzeyi yükseldikçe sıfıra doğru yaklaşmalıdır (RUMM Laboratory Interpreting RUMM2020, part1).

Seçenek cevap eğrisi, seçenek cevap örüntülerinin görsel olarak yorumlanmasını sağlar. Her bir çeldirici için cevap örüntüleri teorik madde karakteristik eğrisi ile karşılaştırılır. Doğru cevabı barındıran seçenek madde karakteristik eğrisini takip eden bir eğri çizer.

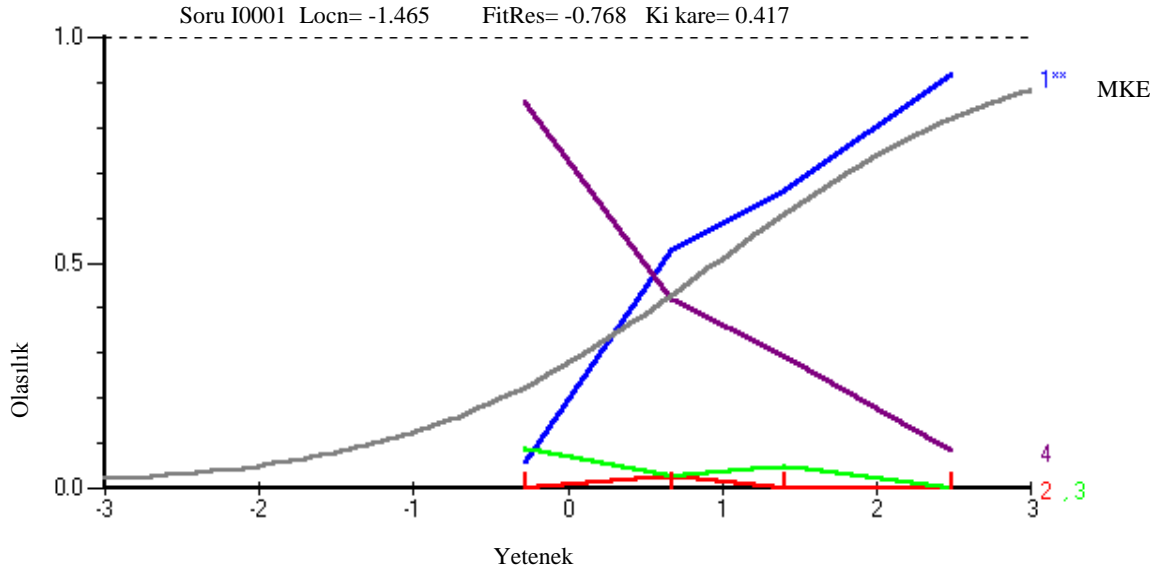
Örnek :

- Madde I0001 (Şekil 7) için madde karakteristik eğrisi ile birlikte dört seçeneğin cevap dağılımları gözlenmektedir. Bu maddenin doğru seçeneğini “seçenek 3” oluşturmaktadır. Bu madde için seçeneklerin genel olarak bireylere çok iyi alternatifler oluşturmadığı görülmektedir. Bu madde bireylere kolay gelmiş ve θ düzeyi -1.465 olan bu maddede bireylerin çoğu doğru seçeneğe yönelmiş, çeldiriciler yeterince çalışmamıştır.



Şekil 7. Madde I0001 Madde Analizi

- Madde I0020 (Şekil 8) çeldirici örüntüsü incelendiğinde doğru cevabı oluşturan seçenek 1'in madde karakteristik eğrisini takip ettiğini ancak çeldiricilerden birini oluşturan "seçenek 4"ün madde karakteristik eğrisi ve doğru seçenek ile ters bir örüntü sergilediği görülmektedir. Bunun anlamı bu çeldiricinin daha düşük yetenek düzeyinde olan bireyleri çekerken, bireylerin yetenek düzeyleri arttıkça doğru seçeneğin seçilme olasılığı artmakta ve çeldirici 4'ün seçilme olasılığı da düşmektedir. Çeldirici olarak hazırlanmış seçenek 4'ün beklendiği şekilde çalıştığı, diğer çeldiricilerin ise testi alan bireyleri kendilerine çekecek güçlü alternatifler olmadığı gözlenmiştir.



Şekil 8. Madde I0020 Madde Analizi (RUMM Laboratory Interpreting RUMM2020, part1)

Bu drumda çeldiricilerin yeterince iyi çalışmadıkları söylenebilir.

1.2.2 Model Uyumunun İncelenmesi

Model uyumu gösteren istatistikler, gözlenen verinin belirlenen modelinin beklentilerine ne derece uygun olduğunu test eder.

Tümel uyum iyiliğinin değerlendirilmesinde (Öztuna, 2008; Pallant ve Tennant, 2007):

- Madde etkileşim istatistiği
- Birey etkileşim istatistiği
- Madde-incelenen özellik düzeyi etkileşim istatistiğinden yararlanılır .

Bunlardan ilk ikisinde, madde ve bireylere ilişkin etkileşim istatistiği (lojit cinsinden) değerleri standart normal dağılım gösteren bir z puanına dönüştürülür. 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 elde edilen ve incelenen özellik düzeyi boyunca değişmezlik özelliğini gösteren “madde-incelenen

özellik düzeyi etkileşim istatistiğidir". İstatistiksel olarak anlamlı bir ki-kare değeri, maddelere verilen yanıtların hiyerarşik sıralamasının incelenen özellik 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 incelendiğinde; artık değerlerinin ± 2.5 arasında olması modele yeterli bir uyum olduğunu gösterir. Ki-kare istatistiğinde ise, modele uyumsuzluk grafik olarak gösterilir. 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 deneğin bulunması 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, yüksek ayırt ediciliğe sahip maddeler eğriden daha dik, düşük ayırt ediciliğe sahip maddeler ise eğriden daha düz bir şekle sahip olacaktır (Öztuna, 2008; Pallant ve Tennant, 2007).

1.2.3 Maddenin Farklı Fonksiyonlaşmasının İncelenmesi

Model uyumunu etkileyebilecek diğer bir durum, MFF'dir

MFF ile ilişkili kavramlar aşağıda açıklanmaktadır:

Bu bağlamda MFF, incelenen özellik bakımından benzer olan ancak cinsiyet, sosyo-ekonomik düzey gibi değişkenler açısından birbirinden farklı alt gruplarda yer alan bireylerin, bir maddeyi doğru cevaplandırma olasılıklarının farklılaşması olarak tanımlanabilir (Hambleton ve diğerleri, 1991; Zumbo 1999).

MFF durumunda maddeler, testi alan alt gruplardan en az biri için daha az geçerliğe sahiptir. Testte yanlış maddelerin bulunması, testin geçerliğini ve buna bağlı olarak güvenilirliğini olumsuz yönde etkiler (Hagquist, 2010).

MFF, tekdüze (uniform) ve tekdüze olmayan (nonuniform) olmak üzere iki grupta değerlendirilir;

Tekdüze MFF: Gruplar arasında ilgili maddeyi doğru yanıtlama olasılığındaki farklılığın, bütün yetenek düzeylerinde tutarlılık göstermesi durumunu belirtir. Diğer bir ifadeyle, ölçme aracındaki herhangi bir madde için bir grup, θ düzeyi boyunca diğerinden anlamlı derecede daha yüksek veya düşük puana sahip ise tekdüze MFF olduğu gözlenir.

Tekdüze olmayan MFF: Gruplar arasında varılan maddeyi doğru yanıtlama olasılığındaki farklılık, bütün yetenek düzeylerinde sabit değilse bu tür MFF tek düze olmayan MFF olarak adlandırılır. Diğer bir deyişle, ölçme aracındaki herhangi bir madde için θ 'nın belli bir değerine kadar bir grup, o değerden sonra da diğer grup anlamlı derecede daha yüksek/düşük puana sahip ise tekdüze olmayan MFF'den bahsedilir (Swaminathan ve Rogers, 1990).

MFF çalışmalarında ilk adım, alt grupların maddeleri yanıtlama yapıları arasında farklılık olup olmadığını, diğer bir deyişle maddelerin farklı fonksiyonlaşıp fonksiyonlaşmadığının belirlenmesine yönelik istatistiksel analizlerin uygulanmasıdır. Bir maddenin yanlı olduğuna karar verebilmek için öncelikle o maddenin farklı fonksiyonlaşması gerekir. Ardından, maddenin farklı fonksiyonlaşmasının olası sebepleri tespit edilerek, maddenin ölçülmek istenen yapıdan bağımsız olarak bazı alt gruplar için adil olmayan bir avantaj sağlayıp sağlamadığının uzman kişiler tarafından belirlenmesi gerekir (Camilli ve Shepard, 1994; Zumbo, 1999).

MFF değerlendirme yöntemleri Klasik Test Kuramına dayalı yöntemler, χ^2 ve Madde Tepki Kuramına dayalı yöntemler olarak sınıflandırılabilir. MTK kullanılarak geliştirilen yöntemler; madde karakteristik eğrileri arasında kalan alan indeksi, madde parametreleri farklılığı indeksi ve olabilirlik oran analizi yöntemi şeklinde sıralanabilir.

MFF olmayan bir test maddesinin farklı örneklemelerden gelen ve aynı yetenek düzeyinde olan bireylerde aynı şekilde işlemesi, aynı madde için

farklı gruplarda yer alan aynı yetenek düzeyindeki bireylerden elde edilen madde karakteristik eğrilerinin benzer olması gerekir (Rodney ve Drasgow, 1990). Bu anlamda, madde yanlılığı analizini MTK modelleriyle yapmak, bir test maddesinin örneğin, kadın ve erkek cinsiyet grupları gibi iki ayrı gruptan elde edilen madde karakteristik eğrilerinin karşılaştırılmasına dayanmaktadır.

Madde karakteristik eğrileri arasındaki farklılık, farklı iki gruptan gelen aynı yetenek düzeyine sahip bireylerin o maddeye doğru cevap verme olasılıkları arasında fark olduğuna işaret eder. Madde karakteristik eğrileri arasındaki farklılık, madde karakteristik eğrilerini tanımlayan parametrelerin veya madde karakteristik eğrileri arasında kalan alanın karşılaştırması gibi farklı yöntemlerle tespit edilebilir (Hambleton ve diğerleri, 1991).

Bu karşılaştırmalarda kullanılabilecek birinci yol madde parametrelerini karşılaştırmaktır. Bu yolda, farklı gruplarda yer alan aynı yetenek düzeyindeki bireyler üzerinden elde edilen madde parametre değerleri aynı ise madde karakteristik eğrilerinin de aynı olması bekleneneğinden madde yansız; madde parametre değerleri arasındaki farklar büyüdükçe madde karakteristik eğrileri de farklılaşacağından madde yanlı yorumu yapılabilir. Karşılaştırmalarda kullanılacak ikinci yol ise farklı gruplardan elde edilen madde karakteristik eğrileri arasında kalan alanların hesaplanmasına dayanır. Madde karakteristik eğrileri arasındaki alan küçükse (sıfıra yakınsa) madde yansız; büyükse madde yanlı yorumu yapılabilir. Madde karakteristik eğrileri arasındaki alanın büyüklüğü sıfırdan uzaklaştıkça (arttıkça) maddenin yanlılığı da artmaktadır (Raju, 1988).

MFF tespit edilen maddelerin testten çıkarılması özellikle sınırlı sayıda maddenin bulunduğu ve ölçülecek özelliğe ilişkin yapının ve kapsam geçerliğinin zarar göreceği durumlarda sorun yaratabilmektedir. Gereksizce çok sayıda maddenin testten çıkarılması sınırlılıkların ortaya çıkmasına neden olabilir. Bu nedenle, MFF'nin gerçekten var olup olmadığı ve yeteneğin kestirilmesinde anlamlı bir fark yaratıp yaratmadığı araştırılırken dikkatli olunmalıdır. Crane, Hart, Gibbons ve Cook, (2006) yapmış oldukları çalışmalarında orta düzeyde MFF tespit edilmiş olmasının bireylerin yetenek kestiriminde pratikte bir farklılık oluşturmadığını göstermişlerdir.

Yapay (artificial) MFF

Bir grup madde, cinsiyet, sosyo-ekonomik düzey gibi incelenen faktörün bir düzeyi için, diğer bir grup madde ise bu faktörün diğer bir düzeyi için daha zor veya kolay başarıyorsa ve bu maddelerden faktörün düzeyi kadar alt test (subtest-testlet) yapıldığında MFF ortadan kalkıyorsa, bu tür MFF, yapay (artificial) MFF olarak adlandırılır.

Bir maddenin bir grup lehine davrandığı gerçek MFF için yapılan değerlendirme süreçleri, diğer maddelerin diğer gruplar lehine davrandığı MFF'e sebep olabilir. Bu durumda ortaya çıkan MFF, MFF belirleme sürecinin sonucunda olan yapay bir MFF'tir. Bu tür MFF'i tanımlamak üzere yapay MFF (artificial MFF) kavramı kullanılmaktadır. Yapay MFF gerçek MFF gösteren maddelerin tanımlanmasında ve yorumlanmasında da yanılgıya yol açabilir (Andrich ve Hagquist, 2011).

Alan yazında, MFF gösteren maddelerden kaynaklanan potansiyel problemlerin iyi tanımlanmış olmasıyla birlikte, MFF'i saptama, MFF'i değerlendirme ve problemi ortadan kaldırma yöntemlerinin hala geliştirilmeye ve detaylandırılmaya ihtiyaç olduğu belirtilmektedir. Özellikle gerçek MFF ve MFF araştırılma sürecinin bir sonucu olarak ortaya çıkan yapay (artificial) MFF'yi birbirinden ayırabilen yöntemler araştırılmalıdır.

MFF'i ortadan kaldırmaya yönelik olarak aşağıdaki stratejiler denebilir:

- Uzman görüşü olarak MFF'in gerekçeleri ortaya konabilir ve ilgili maddeler testte bırakılabilir.
- Test maddesi çıkarılabilir ya da yeniden düzenlenir.
- MFF olan grubun verileri kayıp veri (missing) olarak ele alınabilir.
- Rasch analizi ile gerçek ve yapay MFF'ı birbirinden ayırmak ve yapay MFF sorununu gidermek mümkündür. Hem tekdüze hem de tekdüze olmayan MFF'ye yönelik olarak standartlaştırılmış artık (standardised residuals) değerleriyle varyans analizi (ANOVA) yapılır.

Bu çalışmada, MFF ve Yapay MIF Rasch analizi ile değerlendirilmiştir. Rasch analizi ile MFF varlığı, hem istatistiksel hem de grafik olarak belirlenebilir. Bireylere ait örneğin cinsiyet gibi bir değişkenin her bir düzeyi ve incelenen özelliğin farklı düzeyleri (sınıf aralıkları) için her bir maddenin doğru yanıtlanma olasılığı üzerinden iki yönlü varyans analizi gerçekleştirilir. Tekdüze MFF, bireye ilişkin değişken (cinsiyet) için anlamlı ana etki, tekdüze olmayan MFF ise anlamlı etkileşim terimi (cinsiyet x sınıf aralığı) ile belirlenir (Öztuna, 2008).

1.2.4 Tek Boyutluluğun ve Yerel Bağımsızlığın İncelenmesi

Tek boyutluluğun belirlenmesi için artıklara ilişkin temel bileşenler analizi (TBA) uygulanır. TBA sonuçlarına göre, artıklarda herhangi anlamlı bir yapı 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 gözlenebilir (Pallant ve Tennant, 2007).

Soru bankasının psikometrik özelliklerinin incelenmesi aşamasının sonunda, çeldirici analizinde sorunlu olan, modele uyum göstermeyen, tek boyutluluk varsayımını bozan maddeler soru bankasına alınmamıştır. Sınıf düzeyi ve cinsiyet bakımından MFF gösteren maddeler alan uzmanları ile incelenmiştir.

1.3. Soru Bankasının Kapsam Geçerliği (Content Validity)

Geçerlik, genel anlamda, bir ölçme aracının ya da ölçme yönteminin ölçmeyi amaçladığı özelliği, başka herhangi bir özellikle karıştırmadan, doğru olarak ölçebilme derecesi olarak tanımlanabilir (Baykul, 2000). Bu doğrultuda geçerliği araştırma testin kullanım amacına uygunluk gösteren kanıtları ortaya

koyma süreçlerini içerir. Geçerlik türleri genel olarak üç başlık altında toplanır:

- Kapsam geçerliği
- Ölçüte dayalı geçerlik
- Yapı geçerliği
 - İçsel
 - Dışsal

Kapsam geçerliği: Ölçülmek istenen kapsamla, geçerliği araştırılan testin ölçtüğü kapsamın örtüşme derecesidir. Bir ölçme aracının ölçülmek istenen davranışları, özellikleri ne derece kapsadığının belirlenmesidir. Bir testin, bu testle ölçülmek istenen davranışları ne derecede kapsadığıdır. Örneklem olarak belirlenen test veya ölçek maddeleri, belirli bir amaca yönelik olarak ait olduğu kavramsal evreni temsil etmelidir. Test veya ölçek maddeleri kavramsal evreni temsil ettiği oranda kapsam geçerliliğine sahiptir (Şencan, 2005).

Kapsam geçerliğini, sağlamak için uzman kişiye danışma, belirtke tablosunu dikkate alma ve konu hedef ağırlığına bakmak gerekir. Kapsam geçerliğinin değerlendirilebilmesi için testin kapsamı beklenen davranışların, hedeflerin belirlenmiş olması gerekmektedir.

AÜTF’de, öğrencilerin bilişsel alanda kazanımlarını biçimlendirmeye yönelik değerlendirmek üzere yapılan Gelişim Sınavında maddelerin belirlenmesi için Temel ve Klinik Bilimler, Anabilim Dalları ve Sistem Temelli bir matris, uzman bir grup tarafından oluşturulmuştur. Bu matrisin oluşturulmasında ve ağırlıklandırılması çalışmalarında yurt dışından örnekler (Tıp Lisans Sınavı-United States Medical Licensing Examination-USMLE, Maastricht Üniversitesi gelişim sınavı belirtke tablosu, exam master vb.), Ulusal Çekirdek Eğitim Müfredatı, Uluslararası Hastalık Sınıflaması-10 (ICD-10), Sağlık Bakanlığı sık görülen hastalıklar listesi, Tıpta Uzmanlık Sınavı soru dağılımından yararlanılmıştır.

1.4 Soru Bankasının İç Tutarlılığı / Güvenirlik (Consistency/Reliability)

Oluşturulan soru bankasındaki bulunan 103 maddeden oluşan testin iç tutarlılığının bir tahmini olarak Birey Ayırsama İndeksi (BAİ) [Person Separation Index-(PSI)] ve K-R 20 hesaplanmıştır. 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. KR-20 iç tutarlılık katsayısıdır ve maddeleri 1 ve 0 olarak puanlanan testlerde testin bütününe ait güvenirlilik katsayısı hesaplamak için kullanılmaktadır. KR-20 ile güvenirliliği bulmak için testteki her maddenin güçlük derecesinin bilinmesi zorunludur. Bu nedenle de testin madde analizi yapılmış ve maddelerin güçlük derecesinin bulunmuş olması gerekir (Baykul, 2000; Öztuna, 2008; Pallant ve Tennant, 2007).

2. Simüle uygulama için veri türetilmesi ve simüle verilerle BUT uygulama (Soru Bankalarının Dış Geçerliliği- External Validity)

Geliştirilen soru bankası madde parametreleri kullanılarak; benzetim yoluyla elde edilen veriler üzerinde BUT uygulanmıştır. Bu amaçla SmartCAT paket programı kullanılmıştır.

Bu programda, BUT yönteminin uygulanmasında aşağıdaki aşamalar izlenmiştir:

- ***ilk maddenin seçimi:*** Bireylere 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 zorluk parametresine sahip olan madde başlangıç maddesi olarak seçilmiştir.
- ***İncelenen özellik (θ) düzeyinin tahmini:*** Bu çalışma kapsamında bireylerin θ düzeylerinin kestirimi için “sonsal beklenti” (Expectation a Posteriori, EAP) kestirimi kullanılmıştır.

Kullanılan yetenek kestirimi yöntemine göre ilk maddenin ardından takip edilebilecek farklı kestirim yöntemleri vardır (En Yüksek Olabilirlik-Maximum Likelihood-MLE yöntemi ve Bayesien yöntemler; “Sonsal Beklenti-Expectation a Posteriori-EAP, Maximum Posteriori-MAP). MLE uygulamasında, yetenek kestirimi için bireyin en az bir doğru ve bir yanlış yanıt vermesine ihtiyaç duyulmaktadır. Bu şart sağlanmadığı durumda yetenek kestirimi yapılamaz. Bu bakımdan bilgisayar, sıradaki maddeleri bireyin en az bir doğru ve bir yanlış yanıt vermesini sağlayana kadar seçer. Bir başka yetenek kestirim yöntemi olan Bayesian Expected A Posteriori (EAP) için böyle bir şart yoktur ve yetenek kestirimi ilk maddeye verilen yanıtın ardından hemen başlayabilir. EAP bireyin θ düzeyinin sonsal dağılımının varyansını minimize eden maddelerin seçimi temeline dayalıdır. Bayesçi yaklaşımda, bireyin θ düzeyi hakkında önsel bir varsayımdan yola çıkılarak, sonsal θ düzeyi belirlenmeye çalışılır. EAP iki kategorili puanlanma gerektiren BUT uygulamalarında, iteratif/ardışık süreçlere ihtiyaç duymaması, daha küçük standart hata kestirimi içermesi nedeni daha yaygın olarak kullanılan bir yöntemdir (Wang ve Vispoel, 1998)

- **Soru seçimi:** Bu çalışmada madde seçiminde “maksimum sonsal ağırlandırılmış bilgi kriteri” (Maximum Likelihood Weight Information) kullanılmıştır.
- **Durdurma kuralı:** Literatürde BUT uygulamalarında durdurma kuralı olarak kullanılabilecek kriterlerden bazıları aşağıda sıralanmıştır (Simms ve Clark, 2005; Linacre, 2000);
 - o Sabit madde sayısı
 - o Belirlenmiş standart hata değeri
 - o Yukarıda tanımlanan kuralların bir arada kullanılması

Bu çalışmada durdurma kuralı olarak farklı standart hata değerleri (0.50, 0.40 ve 0.30) kullanılmıştır. Bireylerin θ düzeyleri belirlenen standart hata ile tahmin edildiği durumda BUT, o birey için sonlandırılmıştır.

Çalışmada, BUT simüle uygulama için Rasch analizi sonucunda elde edilen madde parametreleri kullanılarak, 1000 öğrencinin final soru bankası maddelerine verdikleri yanıtlar benzetim programı SmartCAT aracılığıyla türetilmiş ve türetilen bu yanıt yapıları kullanılarak öğrencilerin yetenek düzeyleri (θ_{Rasch}) elde edilmiştir. SmartCAT programı kullanılarak türetilen yanıtlarla aynı 1000 öğrencinin yetenek düzeyi (θ_{BUT}) simüle BUT yöntemiyle de elde edilmiştir.

Benzetim yoluyla türetilen veriler üzerinde gerçekleştirilen BUT uygulaması için ortalaması 0, varyansı 1 olan ve ortalaması 0, varyansı 3 olan iki farklı normal dağılımdan 1000'er birey için yanıt türetilmiştir.

3. Bireylere ilişkin BUT yöntemi ile elde edilen yetenek düzeyleri (θ_{BUT}) ile kağıt kalem testinden elde edilenler (θ_{Rasch}) arasındaki uyumun incelenmesi

BUT uygulamasından elde edilen incelenen özellik düzeylerinin (θ_{BUT}), tüm maddeler kullanılarak Rasch analizinden elde edilenler (θ_{Rasch}) ile karşılaştırılmasında, Spearman'ın korelasyon katsayısı, sınıf içi korelasyon katsayısından (SKK) ve Bland-Altman uyum sınırlarından yararlanılmıştır.

Sınıf-içi korelasyon katsayısı

Aynı yanıtlayıcıdan alınan tekrarlı ölçümler veya aynı yanıtlayıcı üzerinde iki ya da daha fazla sayıda gözlemcinin ölçümleri arasındaki uyumun değerlendirilmesinde sınıf içi korelasyon katsayısı (SKK) kullanılır.

Ölçümler sürekli olduğunda, gözlemci içi veya gözlemciler arası uyumun değerlendirilmesinde, sınıf içi korelasyon katsayısı (SKK) kullanılır.

SKK, hem birim hem de varyansları bakımından aynı sınıfa ait ölçümler arasındaki ilişki miktarını belirlemek amacıyla geliştirilmiştir.

SKK aynı yanıtlayıcıdan elde edilen ölçümlerin değişkenliğini, yanıtlayıcılar ve ölçümler veya gözlemciler üzerinden elde edilen toplam değişkenlik ile karşılaştırılır. Bu açıdan değerlendirildiğinde, SKK, genel tanımı ile iki varyansın birbirine oranıdır.

$$SKK = \frac{\sigma_{DA}^2}{\sigma_{DA}^2 + \sigma_{DI}^2}$$

σ_{DA}^2 : yanıtlayıcılar arası varyansı

σ_{DI}^2 : yanıtlayıcılar içi varyans

Çalışmanın amacı, tasarımı ve ölçümlerin alınma tipine bağlı olarak SKK hesaplanmasında kullanılan farklı formüller mevcuttur. Sadece yanıtlayıcılardan kaynaklanan sistematik bir değişkenlik varsa, tek yönlü modeller; yanıtlayıcıların yanı sıra, gözlemcilerden kaynaklanan bir değişkenlik de varsa iki yönlü modeller kullanılmalıdır. Tüm modeller için denek etkisi rastgele alınır. Bununla beraber gözlemcinin etkisi, modelin tanımlanmasında etkilidir. Eğer gözlemcinin etkisi rastgele ise, model “rastgele etki modeli”, sabit ise, “karma etki modeli” olarak tanımlanır. Model seçiminden sonra, diğer iki önemli ayrım ise, SKK’nın, elde edilen gözlem değerlerinin tek bir ölçüm sonucuna mı yoksa birden çok ölçümün toplamı ya da ortalaması kullanılarak mı hesaplandığı ve SKK hesaplamalarının, “tutarlılık” veya “mutlak uyum” tanımlarından hangisi kullanılarak yapıldığıdır (Ateş, 2009).

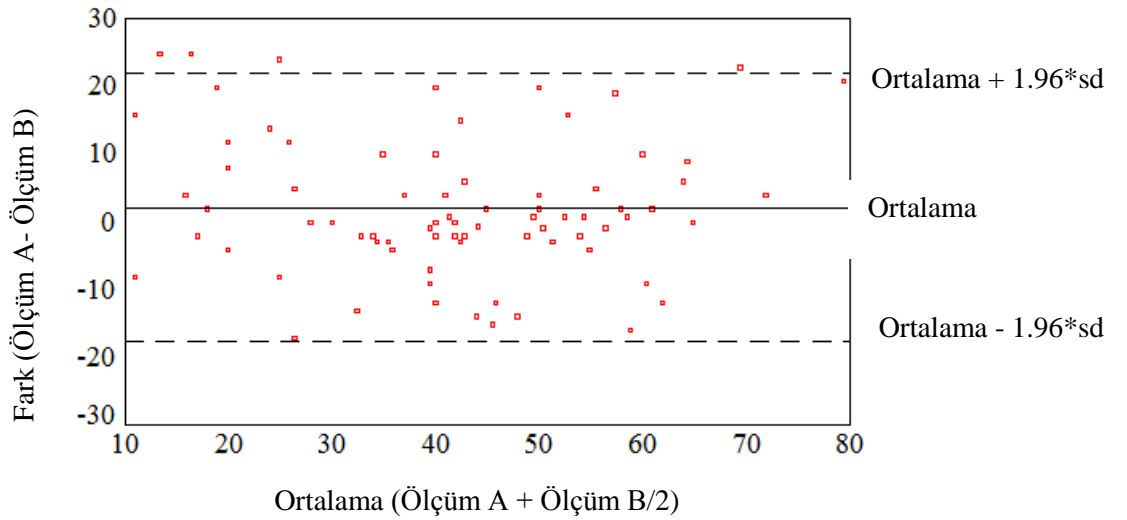
Bu çalışmada SSK hesaplama modeli olarak iki yönlü rastgele etki modeli kullanılarak mutlak uyum tahmini elde edilmiştir (Ateş, 2009).

Bland-Altman Uyum Sınırları

Aynı bireye ilişkin farklı yöntemlerden elde edilen ölçümlerin bulunduğu durumda ölçümler arası uyum araştırılmasında kullanılabilecek alternatif bir analiz yöntemi Blant-Altman tarafından ortaya konulmuştur (Altman ve Bland, 1983). Bu yöntemde öncelikle iki ölçüm arasındaki farklar bulunur, bu farkların ortalaması alınır. Bu ölçüm farklarının standart sapması da

bulunarak ana kütle ortalama farkları için bir güven aralığı hesaplanır. İki yöntemden elde edilen ölçümlerin ortalamaları ve farklarının saçılım grafiği çizilir.

Saçılım grafiğinin yatay eksenini değerlendirmeye alınan bireylerin ölçümlerinin ortalamasını, dikey eksen ise iki ölçüm arasındaki farkı gösterir. Grafikte yatay eksene paralel olarak ölçümlerin ortalaması ve bu ortalamanın üzerinde ve altında yer alan alt ve üst güven aralıkları çizilir (Şekil 9).



Şekil 9. Bland-Altman Uyum Sınırları Grafiği Örneği

Farklara karşı ortalamaların grafiği, ölçüm hataları ile gerçek değerler (ortalamalar) arasında olabilecek herhangi bir ilişkinin incelenmesine olanak sağlar. Her bir birey için ölçümler arası farkların, ortalama ve güven aralığına göre nasıl yerleşmiş oldukları değerlendirilir. Ayrıca bu grafikten sistematik hata (farkların sistematik olarak 0'dan farklı olup olmadığının) ve rastgele hatanın (farkların ne kadar yaygın olduğunun) incelenmesi de mümkündür. Eğer farklar normal dağılıma sahipse farkların ortalamasının etrafında rastgele dağılması ve % 95'nin $\bar{x} \pm 1,96sd$ arasında olması beklenir ve iki yöntemin uyumlu sonuçlar verdiği söylenebilir (Bland ve Altman, 1986; Bland ve Altman, 1999).

Bu çalışmada θ_{Rasch} ve θ_{BUT} değerlerinin uyumları Bland-Altman uyum sınırlar grafikleri ile incelenmiştir.

BÖLÜM IV

BULGULAR VE YORUMLAR

Bu bölümde, Ankara Üniversitesi Tıp Fakültesinde yürütülmekte olan gelişim sınavı için bir soru bankası oluşturulması ve bu soru bankasını kullanarak Bilgisayar Uyarlamalı Test (BUT) yönteminin uygulanabilirliğinin ortaya konulmasını amaçlayan bu araştırmanın bulguları sunulmuştur. Bulguların sunuluşunda araştırmada cevaplandırılması planlanan soruların ele alınış sırası temel alınmış; önce her soru hatırlatılmış ve daha sonra da sorulara ilişkin bulgular sunularak tartışılmıştır.

1. Gelişim Sınavı için tıbbi bilgi düzeyini ölçmek amacı ile geçerli ve güvenilir testler oluşturabilecek bir soru bankası oluşturulabilir mi?

Bu soruya ilişkin bulgular alt başlıklarda verilmiş

1.1. Soru Bankasının psikometrik özellikleri (Soru Bankasının İçsel Yapı Geçerliliği - Internal Construct Validity ve Güvenirlik)

Başlangıç Soru Bankasının Klasik Yöntemle Uygulanması: Çalışma kapsamında öncelikli olarak temel, dahili ve cerrahi tıp bilimleri alanlarını kapsayan 200 maddeden oluşan başlangıç soru bankası, 1206 öğrenciye uygulanmıştır. Bu sınavda öğrenci sayılarının dönemlere göre dağılımı Çizelge 2’de verilmiştir.

Çizelge 2. 2011 Gelişim Sınavına Katılan Öğrenci Sayısının Dönemlere Göre Dağılımı

<i>Dönem</i>	<i>Öğrenci sayısı (N)</i>	<i>Sınava katılan öğrenci sayısı (n)</i>	<i>Sınava katılma yüzdesi (%)</i>
1	314	287	91,4
2	228	202	88,6
3	299	292	97,7
4	267	250	93,6
5	236	175	74,2
Toplam	1344	1206	89,7

Bu aşamada, 200 maddelik soru bankasının psikometrik özellikleri tek parametrelili Rasch modeli kullanılarak incelenmiştir. Buna göre, Rasch analizi sonucunda madde analizinde sorunlu bulunan veya modele uyum göstermeyen maddeler çalışmadan çıkarılmış; BUT yöntemi için kullanılacak olan final soru bankası elde edilmiştir.

1.1.a. Seçenek Analizi

Rasch analizinin ilk aşaması olarak 2011 yılı gelişim sınavında sorulan maddelerin kategorileri için iki sonuçlu Rasch modeli kullanılarak seçenek analizi (distractor analysis) yapılmıştır. Bu analiz sonucunda sorunlu olan 33 madde, alan uzmanları ile yapılan görüşmeler bankasından çıkarılmıştır. Sonuç olarak 200 maddenin 167'si ile analizlere devam edilmiştir.

1.1.b. Model Uyumunun İncelenmesi

İkinci aşamada soru bankasında kalan 167 maddenin model uyumları incelendiğinde, 64 maddenin modele uyum göstermediği belirlenmiştir. Geriye kalan 103 maddenin Bonferroni düzeltmeli olasılık (p) değerleri incelendiğinde modele uyumlu oldukları belirlenmiştir (Çizelge 3). Ayrıca 103 maddelik final soru bankasının tümel uyum iyiliği istatistikleri incelendiğinde

madde uyum istatistiği ortalaması (standart sapması) 0.402 (1.234), birey uyum istatistiği ortalaması (standart sapması) 0.008 (0.893)'dır. Bu değerlerin ortalamasının 0, standart sapmasının 1 değerine yakın olması, soru bankasında yer alan maddelerin ve bireylerin modele yeterli uyumunun göstergesidir. “Madde- θ etkileşim istatistiği” incelendiğinde ki-kare değeri (p değeri) 1049.33 (0.003) olduğundan ve Bonferroni düzeltmesine göre manidar olmadığından bu soru bankası için maddelere verilen yanıtların hiyerarşik sıralamasının θ düzeyi boyunca değişmediği yani değişmezlik özelliğinin sağlandığı sonucuna ulaşılmaktadır.

Tümel uyum istatistiklerine ek olarak, final soru bankasında yer alan her bir madde için artık değerleri ve ki-kare istatistiği cinsinden hesaplanan uyum istatistikleri Çizelge 3’de sunulmuştur.

Çizelge 3. 2011-2012 Eğitim Öğretim Yılı Gelişim Sınavı “Genel Tıp” Soru Bankasının Rasch Modeline Uyumu (n= 1206)

Madde No	b	Standart Hata	Artık	Ki-kare	Serbestlik Derecesi	Olasılık
1 (m1)	0.450	0.059	2.712	4.047	9	0.908
2 (m2)	0.666	0.059	1.835	5.951	9	0.745
3 (m3)	0.161	0.060	-1.078	9.683	9	0.377
4 (m5)	-0.124	0.062	1.751	10.036	9	0.348
5 (m7)	-0.286	0.064	-0.627	5.949	9	0.745
6 (m8)	-0.146	0.063	0.531	3.161	9	0.958
7 (m9)	1.242	0.062	0.439	7.917	9	0.543
8 (m10)	0.988	0.060	1.027	14.087	9	0.119
9 (m12)	0.883	0.059	1.377	7.983	9	0.536
10 (m13)	1.047	0.060	1.967	10.128	9	0.340
11 (m15)	-0.271	0.064	-1.685	12.958	9	0.165
12 (m19)	1.079	0.060	-0.913	10.384	9	0.320
13 (m21)	-2.769	0.154	-0.364	11.385	9	0.250
14 (m22)	0.809	0.059	1.266	4.721	9	0.858
15 (m24)	-0.349	0.065	1.223	15.112	9	0.088
16 (m25)	-0.547	0.068	1.118	16.137	9	0.064
17 (m26)	-1.334	0.085	-0.578	7.916	9	0.543

Çizelge 3. 2011-2012 Eğitim Öğretim Yılı Gelişim Sınavı “Genel Tıp” Soru Bankasının Rasch Modeline Uyumu (n= 1206) (Devamı)

Madde No	b	Standart Hata	Artık	Ki-kare	Serbestlik Derecesi	Olasılık
18 (m30)	-0.673	0.070	0.921	14.898	9	0.094
19 (m31)	0.201	0.060	2.894	10.344	9	0.323
20 (m33)	0.039	0.061	-0.207	10.740	9	0.294
21 (m35)	0.249	0.060	3.432	20.162	9	0.017
22 (m36)	0.944	0.060	2.187	10.416	9	0.318
23 (m37)	-0.281	0.064	2.182	10.786	9	0.291
24 (m39)	0.471	0.059	-1.055	6.963	9	0.641
25 (m41)	-0.221	0.063	0.363	9.704	9	0.375
26 (m43)	-0.747	0.071	0.290	5.509	9	0.788
27 (m44)	0.395	0.059	2.411	13.970	9	0.123
28 (m45)	-1.465	0.089	0.094	9.199	9	0.419
29 (m46)	0.569	0.059	1.739	7.000	9	0.637
30 (m48)	-0.194	0.063	0.568	8.184	9	0.516
31 (m50)	-0.686	0.070	-0.082	6.296	9	0.710
32 (m52)	0.015	0.061	1.565	11.500	9	0.243
33 (m54)	0.450	0.059	0.222	17.966	9	0.036
34 (m55)	-1.791	0.101	-0.563	11.493	9	0.243
35 (m56)	-0.389	0.065	-0.717	8.558	9	0.479
36 (m58)	0.050	0.061	2.614	16.004	9	0.067
37 (m59)	-0.049	0.062	0.130	6.533	9	0.686
38 (m60)	-0.997	0.076	-1.006	8.024	9	0.532
39 (m61)	-1.450	0.089	-0.344	8.329	9	0.501
40 (m62)	-0.279	0.064	-1.013	12.447	9	0.189
41 (m63)	-1.157	0.080	0.307	16.995	9	0.049
42 (m65)	-2.064	0.113	0.238	9.990	9	0.351
43 (m69)	-0.405	0.066	0.030	7.078	9	0.629
44 (m70)	0.257	0.060	-0.040	11.881	9	0.220
45 (m73)	0.204	0.060	2.389	6.662	9	0.672
46 (m76)	1.123	0.061	1.872	10.796	9	0.290
47 (m79)	-0.311	0.064	1.186	5.011	9	0.833
48 (m87)	-0.274	0.064	-0.481	9.212	9	0.418
49 (m88)	-0.940	0.075	-0.706	9.410	9	0.400
50 (m89)	1.415	0.063	0.721	12.853	9	0.169
51 (m92)	-0.150	0.063	-0.029	10.393	9	0.320
52 (m102)	0.167	0.060	-1.330	10.783	9	0.291
53 (m103)	0.725	0.059	-0.572	14.471	9	0.107
54 (m104)	0.283	0.060	-1.840	12.909	9	0.167
55 (m105)	-1.646	0.096	-0.916	18.944	9	0.026

Çizelge 3. 2011-2012 Eğitim Öğretim Yılı Gelişim Sınavı “Genel Tıp” Soru Bankasının Rasch Modeline Uyumu (n= 1206) (Devamı)

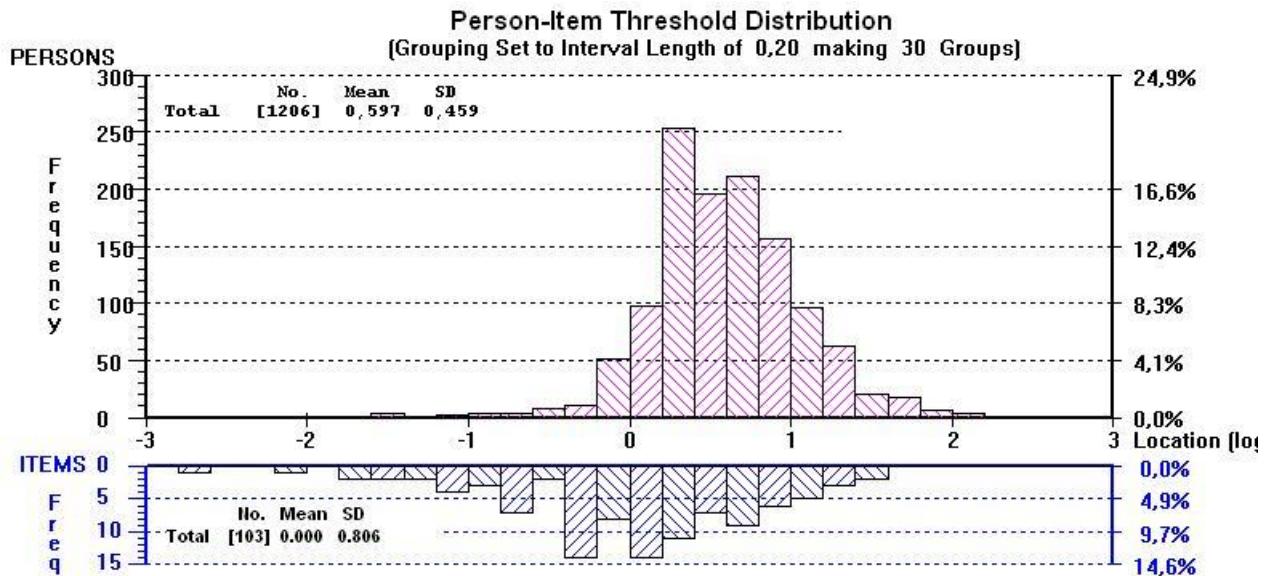
Madde No	b	Standart Hata	Artık	Ki-kare	Serbestlik Derecesi	Olasılık
56 (m106)	1.431	0.064	0.055	4.844	9	0.848
57 (m113)	-0.228	0.063	-0.738	11.141	9	0.266
58 (m114)	0.801	0.059	0.885	5.515	9	0.787
59 (m115)	0.167	0.060	-0.269	6.465	9	0.693
60 (m116)	1.131	0.061	1.534	13.360	9	0.147
61 (m117)	0.198	0.060	1.845	7.156	9	0.621
62 (m119)	0.502	0.059	-1.374	17.268	9	0.045
63 (m121)	0.358	0.059	0.590	5.691	9	0.770
64 (m122)	0.322	0.059	-1.923	18.530	9	0.030
65 (m123)	-1.127	0.080	-1.039	12.310	9	0.196
66 (m124)	-1.177	0.081	-1.308	16.593	9	0.055
67 (m125)	-0.296	0.064	1.033	9.038	9	0.434
68 (m128)	-0.269	0.064	-1.051	10.836	9	0.287
69 (m129)	0.616	0.059	2.739	12.926	9	0.166
70 (m130)	0.561	0.059	0.957	11.488	9	0.244
71 (m131)	-0.765	0.072	-1.099	9.396	9	0.402
72 (m133)	0.743	0.059	-1.062	10.213	9	0.334
73 (m139)	0.191	0.060	1.022	16.195	9	0.063
74 (m140)	-1.369	0.086	-0.807	9.810	9	0.366
75 (m141)	0.879	0.059	1.681	8.408	9	0.494
76 (m142)	1.127	0.061	1.700	7.514	9	0.584
77 (m145)	1.299	0.062	2.395	15.809	9	0.071
78 (m148)	0.687	0.059	0.771	4.412	9	0.882
79 (m151)	-0.695	0.070	-0.107	5.019	9	0.833
80 (m155)	-0.103	0.062	0.011	5.553	9	0.784
81 (m156)	-0.649	0.069	-0.254	14.803	9	0.096
82 (m157)	-0.116	0.062	-0.003	5.754	9	0.764
83 (m158)	0.139	0.060	0.490	10.426	9	0.317
84 (m160)	0.064	0.061	0.884	6.378	9	0.702
85 (m161)	0.718	0.059	2.828	18.111	9	0.034
86 (m165)	-0.278	0.064	0.167	7.022	9	0.635
87 (m166)	0.718	0.059	0.650	13.473	9	0.142
88 (m171)	0.180	0.060	0.138	11.755	9	0.227
89 (m172)	-0.672	0.070	0.355	8.026	9	0.532
90 (m177)	0.329	0.059	1.157	6.059	9	0.734
91 (m178)	0.494	0.059	0.572	6.733	9	0.665
92 (m180)	0.209	0.060	2.800	11.833	9	0.223
93 (m181)	0.653	0.059	-0.988	6.108	9	0.729

Çizelge 3. 2011-2012 Eğitim Öğretim Yılı Gelişim Sınavı “Genel Tıp” Soru Bankasının Rasch Modeline Uyumu (n= 1206) (Devamı)

Madde No	b	Standart Hata	Artık	Ki-kare	Serbestlik Derecesi	Olasılık
94 (m182)	0.104	0.061	0.932	8.226	9	0.512
95 (m183)	0.376	0.059	-1.006	10.950	9	0.279
96 (m185)	1.341	0.063	0.137	14.669	9	0.100
97 (m188)	0.159	0.060	-0.433	10.147	9	0.339
98 (m191)	-0.068	0.062	0.280	2.790	9	0.972
99 (m192)	0.173	0.060	1.213	11.649	9	0.234
100 (m193)	-0.370	0.065	-0.896	14.136	9	0.118
101 (m194)	-1.185	0.081	0.159	6.196	9	0.720
102 (m195)	-0.888	0.074	-0.318	7.108	9	0.626
103 (m200)	0.696	0.059	-1.364	11.517	9	0.242

Çizelge 3’e göre artık değerlerin çoğunun $\pm 2,5$ arasında değerler aldığından ve ki-kare değerlerine ilişkin olasılık değerleri Bonferroni düzeltmeli p değerinden ($0.05/103=0,0005$) daha yüksek olduğundan soru bankasındaki tüm maddelerin modele uyum gösterdiği anlaşılmaktadır.

“Genel tıp” final soru bankasında yer alan maddelerin ve öğrencilerin θ düzeyi boyunca dağılımı Şekil 10’de verilmiştir.



Şekil 10. “Genel Tıp” Final Soru Bankasındaki Maddelerin ve Öğrencilerin θ Düzeyi Boyunca Dağılımı

Soru bankasında yer alan maddelerin b değerlerinin ortalaması ile incelenen grubun θ düzeyleri ortalamasının birbirine uygun olması, dolayısıyla soru bankasının θ 'nın tüm düzeylerinde bilgi verici olması hedeflenir. Soru bankasında bulunan maddelerin farklı yetenek düzeylerine karşılık gelecek uygun dağılım göstermesi beklenir. Bu dağılım izlenmediğinde BUT uygulamalarında yetenek kestirimlerinde standart hata yüksek olacaktır ya da yetenek kestirimlerini istenen doğrulukta yapabilmek için daha uzun testler oluşturmak gerekecektir (Georgiadou ve diğerleri, 2006, Lunz ve Bergstrom, 1994). Şekil 10 incelendiğinde, bireylerin maddelerin konumları tarafından kapsandığı, yani “Genel tıp” final soru bankasının θ 'nın hemen hemen tüm düzeylerinde yeterli sayıda madde içerdiği sonucuna ulaşılmaktadır.

1.1.c. MFF'nin İncelenmesi

Rasch analizinin üçüncü aşamasında, dönem ve kitapçık türü bazında MFF gösteren madde olup olmadığı incelenmiştir. Bu incelemede, kitapçık türü bazında MFF gösteren madde bulunmazken; dönem bazında bazı maddelerin sorunlu olduğu belirlenmiştir. Bu maddeler, alan uzmanları ile beraber detaylı olarak değerlendirildiğinde bir grup maddenin doğru yanıtlanma oranının dönem 1 öğrencileri için, diğer bir grup maddenin doğru yanıtlanma oranının dönem 5 öğrencileri için daha yüksek olduğu görülmüştür. Dönem I öğrencileri tarafından daha kolay yanıtlanan maddelerin, tıp eğitiminin ilk dönemlerinde yer alan temel tıp bilimlerine ait maddeler olduğu belirlenmiş, öğrenciler tarafından daha kolay hatırlanmış olabileceği düşünülmüştür. Dönem V öğrencilerinin daha kolay yanıtladıkları maddelerin ise klinik tıp bilimlerine ait sorular olduğu belirlenmiş ve öğrencilerin bu tür maddeleri yanıtlamada daha deneyimli oldukları fikrine ulaşılmıştır. Bu iki madde grubundan Rasch analizi kapsamında 2 alt test oluşturularak tekrar MFF değerlendirmesi yapıldığında, MFF gösteren maddenin olmadığı belirlenmiştir. Bu tür MFF, literatürde yapay (artificial) MFF olarak tanımlanmakta (Andrich ve Hagquist, 2011) ve bu tür durumlarda hiçbir madde soru bankasından çıkarılmamaktadır.

1.1.d. Tek Boyutluluğun ve Yerel Bağımsızlığın İncelenmesi

Son aşamada, 103 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 herhangi bir yapı gözlenmemiştir ($t=4.6$; güven aralığı %3.4-%5.7). Yerel bağımsızlık varsayımının sağlanıp sağlanmadığını test etmek için artık korelasyon matrisi incelenmiş, 0.30'un üzerinde korelasyona sahip olan madde bulunmamıştır. Çalışmada, soru bankasında yer alan maddelerin tek boyutta toplandığı bu boyutun Gelişim Sınavı için belirlenmiş olan kapsamın tüm alanlarında değerlendirmeye olanak sağlayacağı görülerek bu boyut “genel Tıp” temel boyutu olarak adlandırılmıştır.

1.2. Soru Bankasının Kapsam Geçerliliği (Content Validity)

103 maddelik final soru bankasının içsel yapı geçerliliği Rasch analizi ile incelenmiştir. Buna ek olarak soru bankasının içerik/kapsam geçerliliği alan uzmanları (Temel ,Cerrahi ve Dahili Bilimlerden birer öğretim üyesi ve Tıp Eğitimi Ana Bilim Dalından bir uzman) ile incelendiğinde genel tıp becerisini ölçmek için her bilim alanından Gelişim Sınavı için oluşturulan belirtke tablosunu temsil edebilecek şekilde yeterli sayıda madde içerdiği belirlenmiştir (Ek.2).

Çalışma kapsamında oluşturulan soru bankasının “genel tıp bilgileri” temel alanına ilişkin tek boyutlu bir yapıya sahip olduğu ve maddelerin kapsam dengesini gözeterek bir dağılım gösterdiği gözlenmiştir. Bu nedenle Gelişim Sınavı için BUT uygulamasına olanak vermektedir. Literatürde alt boyutları da olan kapsamlı alanlar için BUT uygulamalarında önemli bir kısıtlılık olarak ortaya konulan soru bankasının tek boyutluluk şartı bu çalışma kapsamında da tartışma alanlarından birini oluşturmaktadır. GS'den önemli bir beklenti, öğrencilere performanslarına ilişkin, hem kapsamın bütünü hem de farklı disiplinler ve alt disiplinler düzeyinde bireyselleştirilmiş geribildirim sağlamak bu yolla öğrencilerin öğrenmelerini ve eğitim

programını desteklemek ve yönlendirmektir. Oluşturulan soru bankası ile öğrencilere toplam bir puan sağlanırken, alt alanlar (organ-sistemler, disiplinler, işlem/görev alanları) özelinde değerlendirme yapılabilmesi için çok daha kapsamlı bir soru bankası veya her bir alt alan için farklı soru bankası geliştirilmelidir (Ackerman ve Davey, 1994; Roex ve Degryse, 2004; van der Vleuten ve diğerleri, 1996a). Literatürde de, birden çok kapsam alanına sahip büyük ölçekli değerlendirmelerde tek boyutluluğu sağlamak için soru bankasının birden fazla alt boyuta ayrılması, her alt boyutun madde bankasındaki madde gücünü dağılımını temsil edecek şekilde madde içermesi ya da kapsamı tüm içerik alanları ile temsil edecek şekilde ağırlıklandırılmış ve güçlük düzeyleri dengelenmiş çok geniş bir soru bankası oluşturma gereksinimi BUT uygulamalarını zorlayıcı bir durum olarak rapor edilmektedir (Bjorner ve diğerleri, 2007; van der Vleuten ve diğerleri, 1996; Thissen ve diğerleri, 2007; Wainer ve diğerleri, 2000; Georgiadou ve diğerleri, 2006; Lunz ve Bergstrom, 1994).

1.3. Soru Bankasının İç Tutarlılık / Güvenirliği (Consistency / Reliability)

Soru bankasında bulunan 103 maddeden oluşan testin/soru bankasının iç tutarlılığı incelendiğinde, hem KR-20 katsayısı ve hem de BAİ 0.77 olarak elde edilmiştir. Bu değerler incelendiğinde 103 maddeden elde edilen testin/soru bankasının güvenilir olduğu ve en az iki grubu birbirinden ayırabildiği sonucuna ulaşılmıştır.

2. İncelenen özellik düzeylerinin (θ) güvenilir olarak kestirilmesinde, Gelişim Sınavının BUT ve kâğıt-kalem testi yöntemi ile uygulanması arasında madde sayısı açısından manidar bir fark var mıdır?

Araştırmada cevap aranmaya çalışılan bu ikinci soruya ilişkin bulgular alt başlık altında ele alınmıştır.

2.1. Simüle uygulama için veri türetilmesi ve simüle verilerle BUT uygulama (Soru Bankalarının Dış Geçerliği- External Validity)

Benzetim çalışmasında, Rasch analizi sonucunda elde edilen madde parametreleri kullanılarak RUMMss programı ile 1000 öğrencinin yanıtları türetilmiştir. Aynı zamanda bu yanıt yapıları ile öğrencilerin yetenek düzeyleri (θ_{Rasch}) de elde edilmiştir. RUMMss programından türetilen yanıtlarla SmartCAT modülü kullanılarak, 1000 öğrencinin yetenek düzeyi (θ_{BUT}) yöntemiyle de bulunmuştur. θ_{Rasch} ve θ_{BUT} değerleri Blant-Altman yöntemi, Spearman ilişki katsayısı ve SKK ile karşılaştırılmıştır.

2.1.1. Simüle BUT uygulamasında $N(0,1)$ için elde edilen sonuçlar

Ortalaması 0, varyansı 1 olan normal dağılımdan 1000 birey için türetilen yanıtlar ile gerçekleştirilen BUT uygulaması sonuçları, farklı standart hata (güvenirlik) değerleri Çizelge 4’de sunulmuştur. Çizelgede yer alan değerler incelendiğinde, 103 maddelik soru bankasının BUT uygulaması sonucunda, güvenilirlik 0.70 olarak alındığında ortalama 11, güvenilirlik 0.75 olarak kabul edildiğinde ortalama 14, güvenilirlik 0.84 olarak alındığında ortalama 24 ve güvenilirlik 0.90 olarak alındığında ise ortalama 45 madde ile kestirim yapılabildiği belirlenmiştir. Her üç koşul için de θ_{Rasch} ile θ_{BUT} değerleri arasında, pozitif yönlü, yüksek, istatistiksel olarak manidar bir ilişki bulunmuştur. θ_{Rasch} ile θ_{BUT} değerleri arasında hesaplanan SKK değerlerinin de oldukça yüksek olması, BUT yöntemi ile elde edilen θ düzeylerinin, tüm maddelerin sorulduğu durumda Rasch analizinden elde edilenlerle uyumlu olduğunu göstermektedir (Çizelge 4).

Çizelge 4: BUT Uygulaması ile Öğrencilerin θ Düzeyi Kestirimleri için Belirlenmiş Standart Hata ve Güvenirlik Düzeyleri için Tanımlayıcı İstatistikler (Ortalaması 0, Varyansı 1 Olan Normal Dağılım İçin)

	BUT uygulamasında kullanılan ortalama (\pmstandart sapma) [ortanca (minimum-maksimum)] madde sayısı	Spearman korelasyon katsayısı	Sınıf içi korelasyon katsayısı [%95 güven aralığı]
Standart hata:0.30 Güvenirlik: 0.90	45 (± 3) [44 (43-50)]	0.975**	0.989 [0.988-0.990]
Standart hata:0.40 Güvenirlik: 0.84	24 (± 3) [23 (23-50)]	0.940**	0.971 [0.967-0.974]
Standart hata:0.50 Güvenirlik: 0.75	14 (± 0.81) [14 (13-21)]	0.886**	0.941 [0.933-0.948]
Standart hata:0.548 Güvenirlik: 0.70	11 (± 0.53) [11 (11-16)]	0.868**	0.928 [0.919-0.937]

** p<0.001

Rasch analizi ve BUT uygulamasından elde edilen yetenek düzeylerinin uyumu Blant-Altman yöntemi ile de değerlendirilmiş; uyum sınırları 0.30, 0.40, 0.50 ve 0.548 standart hata değerleri için ayrı ayrı hesaplanmıştır.

Standart Hata=0.30

θ_{Rasch} ve θ_{BUT} değerlerine ilişkin Bland-Altman uyum sınırları (-0.39, 0.39)'dur. Uyum sınırları incelendiğinde Rasch analizi ile elde edilen yetenek düzeylerinin, BUT yöntemi ile elde edilenlerden 0.39 birim yüksek, 0.39 birim düşük bulunabileceği yorumuna ulaşılır. Buna göre 1000 öğrencinin 942'si (%94.2) uyum sınırları içerisinde kalırken; 58 öğrencinin (%5.8) 34'ünde BUT yöntemi ile elde edilen; 24'ünde Rasch analizi ile elde edilen tıp bilgisi düzeyi daha yüksek bulunmuştur. Final soru bankası için 0.3 standart hata değerinde uyum sınırlarına ilişkin Bland-Altman grafiği Şekil 11(a)'da sunulmuştur.

Standart Hata=0.40

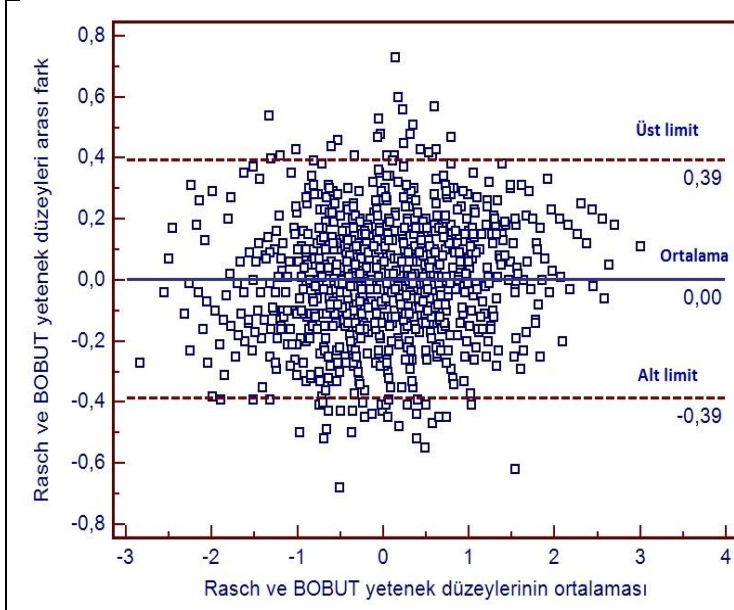
θ_{Rasch} ve θ_{BUT} değerlerine ilişkin Bland-Altman uyum sınırları (-0.64, 0.65)'dir. Uyum sınırları incelendiğinde Rasch analizi ile elde edilen yetenek düzeylerinin, BUT yöntemi ile elde edilenlerden 0.65 birim yüksek, 0.64 birim düşük bulunabileceği yorumuna ulaşılır. Buna göre 1000 öğrencinin 953'ü (%95.3) uyum sınırları içerisinde kalırken; 47 öğrencinin (%4.7) 23'ünde BUT yöntemi ile elde edilen; 24'ünde Rasch analizi ile elde edilen tıp bilgisi düzeyi daha yüksek bulunmuştur. Final soru bankası için 0.4 standart hata değerinde uyum sınırlarına ilişkin Bland-Altman grafiği Şekil 11(b)'de sunulmuştur.

Standart Hata=0.50

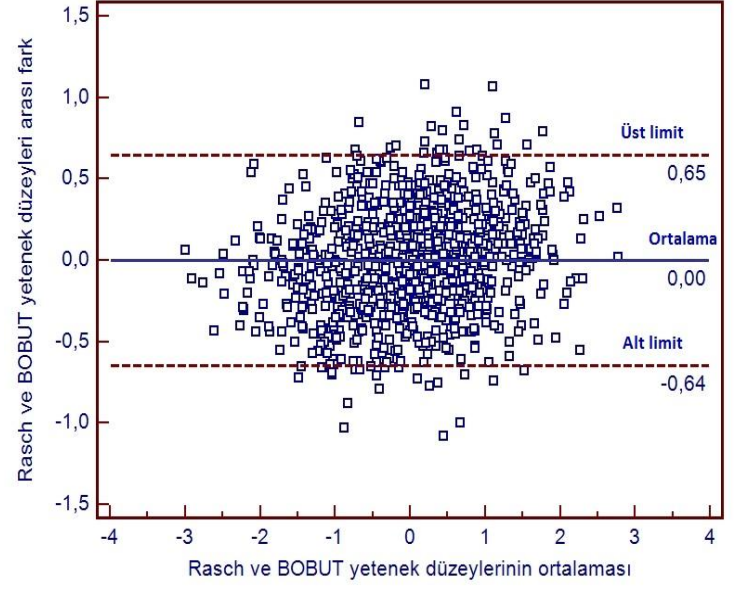
θ_{Rasch} ve θ_{BUT} değerlerine ilişkin Bland-Altman uyum sınırları (-0.83, 0.83)'dir. Uyum sınırları incelendiğinde Rasch analizi ile elde edilen yetenek düzeylerinin, BUT yöntemi ile elde edilenlerden 0.83 birim yüksek, 0.83 birim düşük bulunabileceği yorumuna ulaşılır. Buna göre 1000 öğrencinin 960'ı (%96) uyum sınırları içerisinde kalırken; 40 öğrencinin (%4) 27'sinde BUT yöntemi ile elde edilen; 13'ünde Rasch analizi ile elde edilen tıp bilgisi düzeyi daha yüksek bulunmuştur. Final soru bankası için 0.5 standart hata değerinde uyum sınırlarına ilişkin Bland-Altman grafiği Şekil 11(c)'de verilmiştir.

Standart Hata=0.548

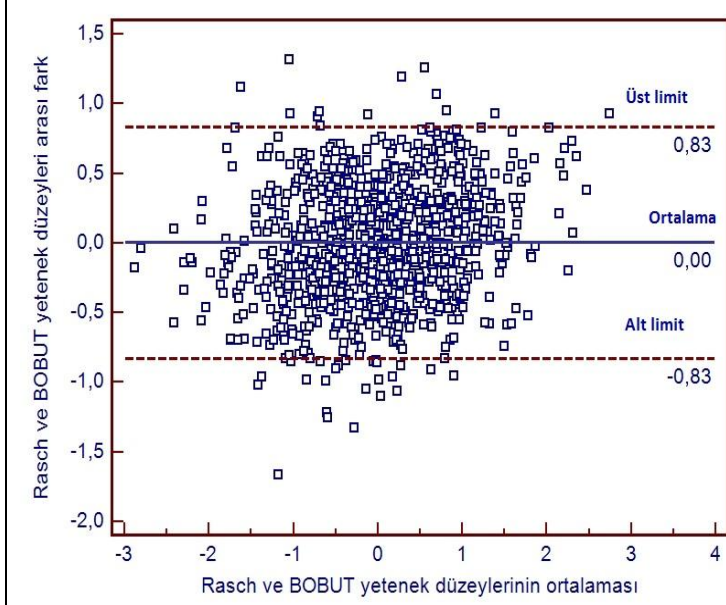
θ_{Rasch} ve θ_{BUT} değerlerine ilişkin Bland-Altman uyum sınırları (-0.92, 0.96)'dir. Uyum sınırları incelendiğinde Rasch analizi ile elde edilen yetenek düzeylerinin, BUT yöntemi ile elde edilenlerden 0.96 birim yüksek, 0.92 birim düşük bulunabileceği yorumuna ulaşılır. Buna göre 1000 öğrencinin 935'i (%93.5) uyum sınırları içerisinde kalırken; 65 öğrencinin (%6.5) 38'inde BUT yöntemi ile elde edilen; 27'sinde Rasch analizi ile elde edilen tıp bilgisi düzeyi daha yüksek bulunmuştur. Final soru bankası için 0.5477 standart hata değerinde uyum sınırlarına ilişkin Bland-Altman grafiği Şekil 11(d)'de verilmiştir.



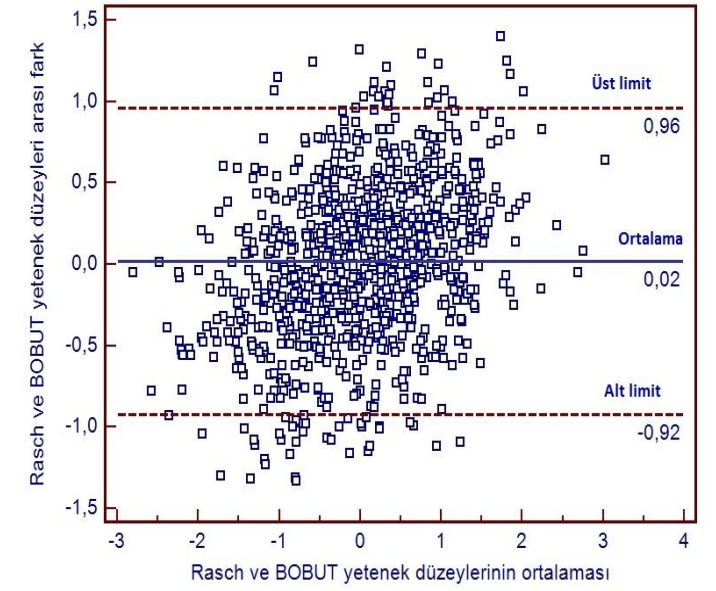
(a)



(b)



(c)



(d)

Şekil 11. Ortalaması 0, Varyansı 1 Olan Normal Dağılım İçin Bland-Altman Grafikleri; (a) SH=0.30, (b) SH=0.40, (c) SH=0.50, (d) SH=0.5478

2.1.2. Simüle BUT uygulamasında $N(0,3)$ için elde edilen sonuçlar

Ortalaması 0, varyansı 3 olan normal dağılımdan 1000 birey için türetilen yanıtlar ile gerçekleştirilen BUT uygulaması sonuçları, farklı standart hata (güvenirlik) değerleri Çizelge 5’de sunulmuştur. Çizelge 5’de yer alan değerler incelendiğinde, 103 maddelik soru bankasının BUT uygulaması sonucunda, güvenilirlik 0.70 olarak alındığında ortalama 12, güvenilirlik 0.75 olarak kabul edildiğinde ortalama 17, güvenilirlik 0.84 olarak alındığında 35 ve güvenilirlik 0.90 olarak alındığında ise 75 madde ile kestirim yapılabildiği belirlenmiştir. Her üç koşul için de θ_{Rasch} ile θ_{BUT} değerleri arasında, pozitif yönlü, yüksek, istatistiksel olarak manidar bir ilişki bulunmuştur. θ_{Rasch} ile θ_{BUT} değerleri arasında hesaplanan SKK değerlerinin de oldukça yüksek olması, BUT yöntemi ile elde edilen θ düzeylerinin, tüm maddelerin sorulduğu durumda Rasch analizinden elde edilenlerle uyumlu olduğunu göstermektedir.

Çizelge 5. BUT Uygulaması ile Öğrencilerin θ Düzeyi Kestirimleri İçin Belirlenmiş Standart Hata ve Güvenirlik Düzeyleri için Tanımlayıcı İstatistikler (Ortalaması 0, Varyansı 3 olan Normal Dağılım)

	BUT uygulamasında ortalama (\pm standart sapma) [ortanca (minimum- maksimum)] madde sayısı	Spearman korelasyon katsayısı	Sınıf içi korelasyon katsayısı [%95 güven oaralığı]
Standart hata:0.30 Güvenirlik: 0.90	75 (\pm 27) [76 (42-103)]	0.998**	0.999 [0.999-0.999]
Standart hata:0.40 Güvenirlik: 0.84	35 (\pm 15) [27 (22-68)]	0.992**	0.995 [0.995-0.996]
Standart hata:0.50 Güvenirlik: 0.75	17 (\pm 3) [15 (15-23)]	0.984**	0.986 [0.984-0.987]
Standart hata:0.548 Güvenirlik: 0.70	12 (\pm 2) [12 (11-16)]	0.975**	0.973 [0.969-0.976]

**p<0.001

Rasch analizi ve BUT uygulamasından elde edilen yetenek düzeylerinin uyumu Bland-Altman yöntemi ile de değerlendirilmiş; uyum sınırları 0.30, 0.40, 0.50 ve 0.548 standart hata değerleri için ayrı ayrı hesaplanmıştır.

Standart Hata=0.30

θ_{Rasch} ve θ_{BUT} değerlerine ilişkin Bland-Altman uyum sınırları (-0.27, 0.28)'dir. Uyum sınırları incelendiğinde Rasch analizi ile elde edilen yetenek düzeylerinin, BUT yöntemi ile elde edilenlerden 0.28 birim yüksek, 0.27 birim düşük bulunabileceği yorumuna ulaşılır. Buna göre 1000 öğrencinin 921'i (%92.1) uyum sınırları içerisinde kalırken; 79 öğrencinin (%7.9) 41'inde BUT yöntemi ile elde edilen; 38'inde Rasch analizi ile elde edilen tıp bilgisi düzeyi daha yüksek bulunmuştur. Final soru bankası için 0.3 standart hata değerinde uyum sınırlarına ilişkin Bland-Altman grafiği Şekil 12(a)'da sunulmuştur.

Standart Hata=0.40

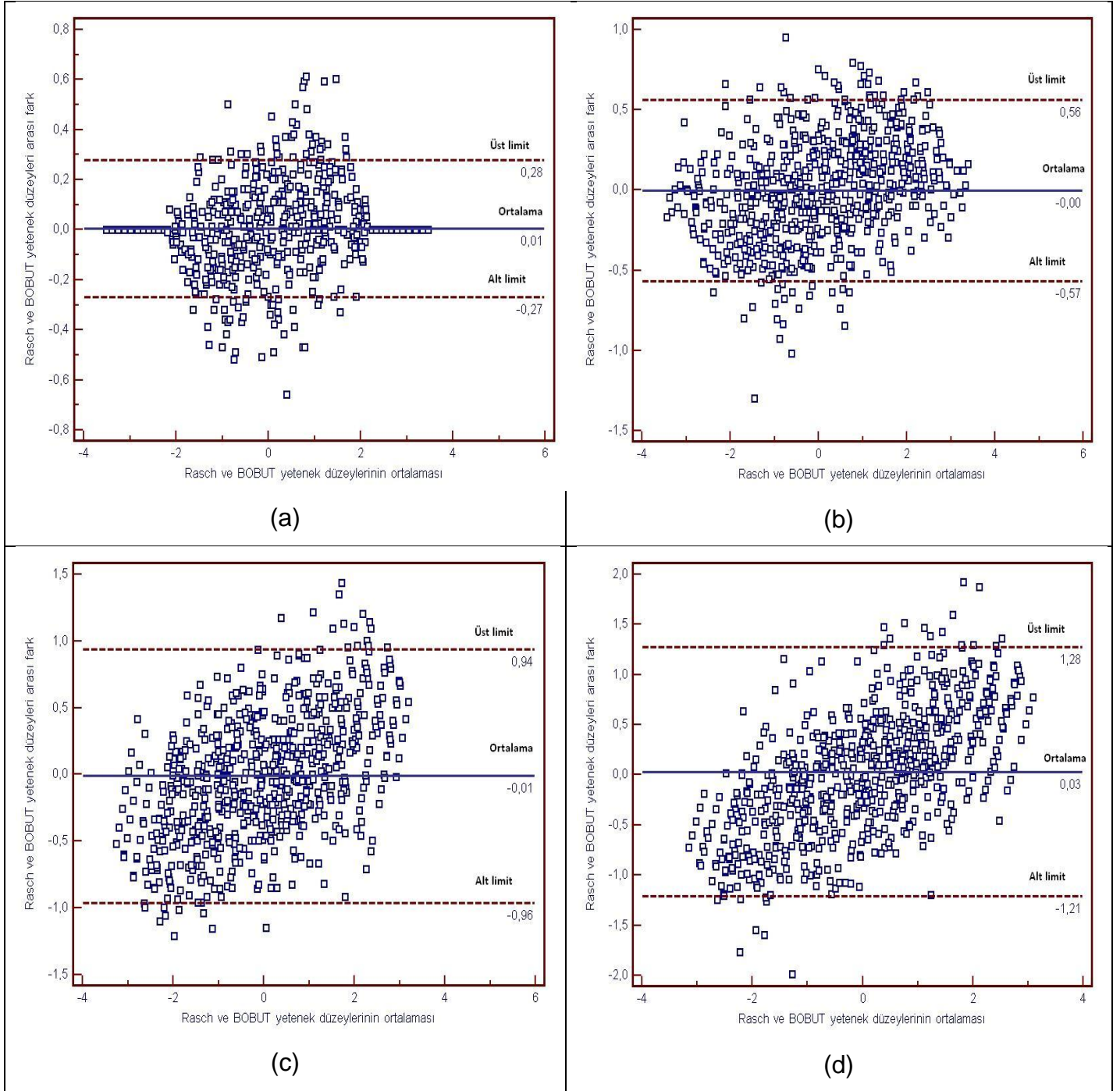
θ_{Rasch} ve θ_{BUT} değerlerine ilişkin Bland-Altman uyum sınırları (-0.57, 0.56)'dır. Uyum sınırları incelendiğinde Rasch analizi ile elde edilen yetenek düzeylerinin, BUT yöntemi ile elde edilenlerden 0.56 birim yüksek, 0.57 birim düşük bulunabileceği yorumuna ulaşılır. Buna göre 1000 öğrencinin 948'i (%94.8) uyum sınırları içerisinde kalırken; 52 öğrencinin (%5.2) 23'ünde BUT yöntemi ile elde edilen; 29'unda Rasch analizi ile elde edilen tıp bilgisi düzeyi daha yüksek bulunmuştur. Final soru bankası için 0.4 standart hata değerinde uyum sınırlarına ilişkin Bland-Altman grafiği Şekil 12 (b)'de sunulmuştur.

Standart Hata=0.50

θ_{Rasch} ve θ_{BUT} değerlerine ilişkin Bland-Altman uyum sınırları (-0.96, 0.94)'dir. Uyum sınırları incelendiğinde Rasch analizi ile elde edilen yetenek düzeylerinin, BUT yöntemi ile elde edilenlerden 0.94 birim yüksek, 0.96 birim düşük bulunabileceği yorumuna ulaşılır. Buna göre 1000 öğrencinin 973'ü (%97.3) uyum sınırları içerisinde kalırken; 27 öğrencinin (%2.7) 10'unda BUT yöntemi ile elde edilen; 17'sinde Rasch analizi ile elde edilen tıp bilgisi düzeyi daha yüksek bulunmuştur. Final soru bankası için 0.5 standart hata değerinde uyum sınırlarına ilişkin Bland-Altman grafiği Şekil 12(c)'de verilmiştir.

Standart Hata=0.548

θ_{Rasch} ve θ_{BUT} değerlerine ilişkin Bland-Altman uyum sınırları (-1.21, 1.28)'dir. Uyum sınırları incelendiğinde Rasch analizi ile elde edilen yetenek düzeylerinin, BUT yöntemi ile elde edilenlerden 1.28 birim yüksek, 1.21 birim düşük bulunabileceği yorumuna ulaşılır. Buna göre 1000 öğrencinin 978'i (%97.8) uyum sınırları içerisinde kalırken; 22 öğrencinin (%2.2) 8'inde BUT yöntemi ile elde edilen; 14'ünde Rasch analizi ile elde edilen tıp bilgisi düzeyi daha yüksek bulunmuştur. Final soru bankası için 0.5477 standart hata değerinde uyum sınırlarına ilişkin Bland-Altman grafiği Şekil 12(d)'de verilmiştir.



Şekil 12. Ortalaması 0, Varyansı 3 Olan Normal Dağılım İçin Bland-Altman Grafikleri; (a) SH=0.30, (b) SH=0.40, (c) SH=0.50, (d) SH=0.548

Çalışmanın bulguları değerlendirildiğinde, BUT yöntemi ile öğrencilerin yetenek düzeylerinin daha az sayıda madde ile kağıt kalem testine benzer şekilde kestirilebildiği görülmüştür. Ortalaması 0, varyansı 1 olan normal dağılım gösteren bir grup için bulgular; Gelişim Sınavının kağıt kalem testi

uygulamaları ve BUT uygulamasının 0.30, 0.40, 0.50, 0.548 standart hata değerleri ile elde edilen yetenek kestirimleri arasındaki korelasyonun yüksek olduğunu (sırası ile 0.98, 0.94, 0.89 ve 0.87) göstermektedir. BUT’da kullanılan madde sayısı standart hatanın farklı düzeyleri için 11 ile 45 arasında değişmektedir. 0.50 standart hata ve 0.75 güvenilirlik düzeyleri için BUT uygulamalarında ortalama 14 madde ile kestirim yapılabilir. BUT ve kağıt kalem testi uygulamalarından elde edilen yetenek kestirimleri arasındaki ilişki (Spearman korelasyon katsayısı: 0.89, Sınıf içi korelasyon katsayısı: 0.94 ve Bland Altman: %96 uyum sınırları içinde) yüksek bulunurken kağıt kalem uygulamalarına göre madde sayısında %89 azalma sağlandığı gösterilmiştir. Güvenirlik düzeyinin 0.90 olduğu durumda madde sayısı %58, güvenilirlik düzeyi 0.70 olduğu durumda da %92 azalmaktadır.

Ortalaması 0, varyansı 3 olan normal dağılım gösteren daha heterojen bir grup için bulgular; Gelişim Sınavının kağıt kalem testi uygulamaları ve BUT uygulamasının 0.30, 0.40, 0.50, 0.548 standart hata değerleri ile elde edilen yetenek kestirimleri arasındaki korelasyon katsayılarının sırası ile 0.98, 0.94, 0.89 ve 0.87 olduğu bulunmuştur. BUT’da kullanılan madde sayısı standart hatanın farklı düzeyleri için 12 ile 75 arasında değişmektedir. 0.50 standart hata ve 0.75 güvenilirlik düzeyleri için BUT uygulamalarında ortalama 17 madde ile kestirim yapılabilir. BUT ve kağıt kalem testi uygulamalarından elde edilen yetenek kestirimleri arasındaki ilişki (Spearman korelasyon katsayısı: 0.98, Sınıf içi korelasyon katsayısı: 0.99 ve Bland Altman: %97 uyum sınırları içinde) yüksek bulunurken kağıt kalem uygulamalarına göre madde sayısında % 83,5 azalma sağlandığı gösterilmiştir. Güvenirlik düzeyinin 0.90 olduğu durumda madde sayısı %27.2, güvenilirlik düzeyi 0.70 olduğu durumda da % 88,3 azalmaktadır.

Weiss (2011) geleneksel testlerle ulaşılan doğruluğa BUT’la %50 oranında daha az madde ile ulaşıldığını, sağlık alanında yapılan çalışmalarda BUT uygulamaları ile test uzunluğunda ortalama %95 azalma sağlandığını belirtmektedir. İşeri (2002) Ortaöğretim Kurumları Seçme ve Yerleştirme Sınavı ve Özel Okullar Sınavının matematik bölümü için yapmış olduğu BUT uygulamalarında, yetenek düzeylerinin daha az madde kullanarak güvenilir

biçimde kestirildiğini göstermiştir. Kalender (2009 ÖSS fen alt testi ile yaptığı çalışmasında 0.3 standart hata değeri kullanılarak BUT uygulamasıyla elde edilen yetenek kestirimleriyle kağıt kalem testi ile yapılan gerçek uygulamalardan elde edilen kestirimler arasındaki korelasyonu %95, uygulama için gerekli olan madde sayısını da 17 bulmuştur. Zitny (2011) BUT uygulamalarına ilişkin 15 araştırmayı inceledikleri çalışmaları sonucunda maddelerin tamamının uygulanması ve BUT uygulamaları ile elde edilen puanlar arasındaki korelasyonları 0.83-0.99 arasında bulmuşlardır. Zitny (2012) ortaokul öğrencilerine uyguladıkları TIP ve VMT testleri ile uyguladıkları çalışmalarında BUT uygulamaları ile TIP’da %55 ve VMT’de %54 daha az madde ile uygulamanın yapılabildiğini bildirmişlerdir. Benzer sonuçlar Bulut’un (2012) yılında yaptığı çalışmalarda da görülmektedir. Bu çalışmada da ALES kağıt kalem testi ve BUT uygulama sonuçlarını farklı standart hata düzeylerini dikkate alarak karşılaştırmışlar ve 0.25, 0.30, 0.40 standart hata eşik değerleri ile BUT uygulanabileceğini, geleneksel kağıt kalem testler ve BUT uygulamalarından kestirilen yetenek düzeyleri arasındaki korelasyonun yüksek olduğunu BUT uygulamaları ile madde sayısında %70 azalma olduğunu rapor etmişlerdir. Kreiter(1999) iç hastalıkları dersi kapsamında yaptıkları çalışmalarında, dersin değerlendirilmesinde kullanılmak üzere 200 maddeden oluşan testi kağıt kalem formatında uygulamışlar ve simüle BUT uygulaması ile elde ettikleri sonuçları karşılaştırmışlardır. 0.75 güvenirlik düzeyi için 52 madde ile yetenek kestirimleri yaptıklarını rapor etmişlerdir.

AÜTF’de uygulanan gelişim sınavı için yapılan çalışmanın bulguları alan yazındaki; BUT uygulamalarının kağıt kalem testi uygulamalarına göre daha az madde ile bireyin yetenek düzeyini güvenilir olarak kestirildiğini gösteren çalışmalarla uyumlu görülmektedir.

BÖLÜM V

SONUÇ VE ÖNERİLER

Sonuçlar

Gelişim Sınavı öğrencilerin, eğitim programı sonunda ulaşmaları beklenen hedeflerle ilişkin kazanımlarına yönelik, geniş kapsamlı bireyselleştirilmiş geribildirimler sağlayan geniş ölçekli bir sınavdır. Gelişim Sınavı öğrencilere ve eğitimcilere öğrenme hedeflerine ulaşmadaki gelişime ilişkin sağladığı bu geribildirimler öğrenmeyi süreç içinde desteklemeyi ve geliştirmeyi amaçlar. Gelişim Sınavı eğitimin sonunda öğrencilerin elde etmeleri beklenen tüm bilişsel çekirdek hedefleri kapsayacak, geçerli ve güvenilir ölçme yapabilecek yeterlikte maddelerden oluşan bir sınav olarak uygulanmalıdır. Paralel testlerle düzenli olarak tekrarlanan bu sınavlar için uygun, geçerli ve güvenilir soru bankasının oluşturulması, öğrenciler ve uygulayıcılar açısından en az iş yükü ile yüksek güvenilirlikli değerlendirmelerin yapılması açısından önemlidir. Bu çalışmada Gelişim Sınavı için soru bankası oluşturmuş ve bu soru bankasının BUT uygulamalarındaki performansı değerlendirilmiştir.

Araştırmanın sonuçları, Rasch yöntemi ile Gelişim Sınavı için uygun bir soru bankası hazırlanabileceği, bu soru bankası kullanılarak Gelişim Sınavının, BUT yöntemi ile uygulanabileceği ve güvenilir sonuçlar elde edilebileceğini göstermektedir. BUT uygulaması ile sınava giren kişilerin sınavın klasik formunda bulunandan çok daha az sayıda madde ile yetenek kestirimleri yapılabileceği sonucuna ulaşılmıştır.

Soru bankasının oluşturulmasında iki kategorili yanıtlar için Rasch modeli kullanılmıştır. Çalışmada soru bankasının Rasch analizi ile

psikometrik özelliklerinin incelenme aşamasında; seçenek analizinde sorunlu bulunan ve modele uyum göstermeyen maddeler çalışmadan çıkarılmıştır ve 103 maddeden oluşan tek bir boyut içeren soru bankası elde edilmiştir. Soru bankasında yer alan, Gelişim Sınavı için belirlenmiş olan kapsamın tüm alanlarında değerlendirmeye olanak sağlayacak şekilde sistem alanlarına dağılım gösteren maddeler “Genel Tıp” olarak adlandırılan tek boyutta yer almıştır. Bu 103 maddeden oluşan soru bankasında yerel bağımsızlığı bozan maddeye rastlanmamıştır.

Soru bankasında MFF gösteren madde olup olmadığı incelendiğinde, dönem bazında MFF gösteren maddeler olduğu görülmüş. Yanlılık gösteren maddeleri iki gruba ayrılarak yapılan Rasch analizi sonucunda MFF gösteren maddenin kalmadığı görülmüş bunun yapay MFF (artificial DIF) olduğuna karar verilmiştir. Soru bankasında bulunan 103 maddeden oluşan testin/soru bankasının BAI ve KR -20 güvenirlik katsayısı değerleri kabul edilebilir sınırlarda bulunmuş (her ikisi için 0.77), 103 maddeden elde edilen testin/soru bankasının öğrencilerin yetenek kestirimlerinde güvenilir olacağı belirlenmiştir. Soru bankasında yer alan maddelerin ve öğrencilerin θ düzeyi boyunca dağılımı incelendiğinde θ 'nın her düzeyi için yeterli sayıda madde olduğu sonucuna ulaşılmıştır.

Çalışma bulguları, Gelişim Sınavı kağıt kalem testi uygulamaları ve BUT uygulamasının ortalaması 0, varyansı 1 olan normal dağılım için 0.30, 0.40, 0.50, 0.548 standart hata değerleri ile elde edilen yetenek kestirimleri arasındaki korelasyonun yüksek olduğunu (sırası ile 0.98, 0.94, 0.89 ve 0.87) göstermektedir. BUT'da kullanılan madde sayısı ortalaması standart hatanın farklı düzeyleri için 11 ile 45 arasında değişmektedir. BUT uygulamalarında ortalama 14 madde ile kestirim yapılabilmektedir. BUT uygulamalarında, kağıt kalem uygulamalarına göre madde sayısında %56.3-86.4 arasında azalma sağlandığı gösterilmiştir. GS kağıt kalem testi uygulamaları ve BUT uygulamasının ortalaması 0, varyansı 1 olan normal dağılım için 0.30, 0.40, 0.50, 0.548 standart hata değerleri ile elde edilen yetenek kestirimleri arasındaki korelasyon katsayıları sırası ile 0.98, 0.94, 0.89 ve 0.87'dir. BUT'da kullanılan madde sayısı standart hatanın farklı düzeyleri için 12 ile

75 arasında değişmektedir. BUT uygulamalarında, kağıt kalem uygulamalarına göre madde sayısında %27.6-88.3 arasında azalma sağlanabildiği ortaya çıkmıştır.

Özet olarak, Gelişim Sınavı için geliştirilen soru bankasının psikometrik özelliklerinin yeterli olduğu, oluşturulan soru bankası ile BUT uygulanabileceği sonucuna ulaşılmıştır. Alan yazında, gelişim sınavının yetenek ölçeğinin alt ucunda bulunan birinci sınıf öğrencileri için zorlayıcı olduğu, bu düzeydeki öğrenciler arasında farklılıkların ortaya konulmasında testlerin daha az duyarlı olabileceği; uzun testlerin yorgunluğa sebep olması ve bunun dikkat ve motivasyonu azaltması, ardışık paralel testlerde madde güçlüğünün benzer olarak sağlanamamasına ilişkin sınırlılıkları olduğu rapor edilmektedir. BUT uygulamaları bu sınırlılıkları en aza indirecek iyi bir çözüm oluşturmaktadır. BUT uygulamaları, kağıt-kalem testi olarak yapılan uygulamalardan farklı olarak öğrenciler kendi yetenek düzeylerine uygun maddelerle değerlendirilebilir ve bu yolla da güvenilir puanlar elde edilirken, yorgunluk, zor/kolay maddeler nedeniyle motivasyon düşmesini de en aza indirebilir. BUT uygulamaları ile maddelerin çoğaltılması, dağıtılması esnasında yaşanması mümkün olabilecek güvenlik sorunlarını da aşmak mümkün olabilecektir.

Tıp eğitiminde Gelişim Sınavında BUT uygulamalarına ilişkin yayınlanmış bir çalışma bulunmamaktadır. Çalışmanın sonucunda tıp eğitiminde Gelişim Sınavında BUT uygulamaları için soru bankasının hazırlanmasının olanaklı olduğu görülmüştür. BUT uygulamalarının avantajları, Gelişim Sınavı uygulamaları içinde önemli katkılar sağlayabilecektir.

İleri Araştırma Önerileri

Çalışmanın bulguları doğrultusunda şu öneriler geliştirilmiştir:

- Bu çalışmada Gelişim Sınavında BUT'un uygulanabilirliği benzetim çalışması ile ortaya konmuştur. Çalışmanın gerçek BUT uygulamaları ile tekrarlanması, öğrencilerin BUT uygulamalarına yönelik, tutumları, davranışları, test ortamının etkileri, bilgisayar kullanıma ilişkin beceriler gibi konuları da değerlendirmeye imkan verecektir.
- Soru bankasında 0'nın tüm düzeyleri için çok sayıda madde olduğunda BUT uygulamasından elde edilen sonuçların güvenilirliği arttacağından Gelişim Sınavı için daha çok soru içeren bir soru bankası oluşturmak için benzer çalışmalar yürütülebilir.
- Soru bankasına madde sağlamanın bir yolu, yeni maddeler yazmak olabileceği gibi, sık kullanılan diğer bir yöntem de kağıt kalem testlerinde uygulanan maddelerin, soru bankası oluşturulmak üzere kullanılmasıdır. Ankara Üniversitesi Tıp Fakültesi'nde Gelişim Sınavı, 2007 yılından bu yana düzenli olarak yapılmaktadır ve her sınav 200 madde içermektedir. Bütün bu uygulamalardaki maddelerin bir soru bankası altında toplanabilmesi için madde kalibrasyonları ve test eşitleme çalışmalarının yapılması önemli katkılar sağlayabilir.
- Gelişim Sınavının temel amacı öğrencilere, hem kapsamın bütünü hem de farklı disiplinler ve alt disiplinler düzeyinde bireyselleştirilmiş geribildirim sağlamaktır. Soru bankası ile öğrencilere toplam bir puan sağlamanın yanı sıra, alt alanlar (organ-sistemler, disiplinler, işlem/görev alanları özelinde) için değerlendirme yapılabilmesi ve olgu özerkliğine (case-specificity) ilişkin kısıtlıkların ortadan kaldırılabilmesi için çok daha kapsamlı ve/veya her bir alt alan için farklı soru bankası geliştirilmesine yönelik çalışmalar önerilebilir.

- Tıp eğitiminde yapılan deęerlendirmelerin farklı yetenek, bilgi alanlarını içermesinden ve olgu özgüllüğünden kaynaklanan kısıtlılıkların giderilmesine yönelik olarak farklı alt alanları kapsayan deęerlendirmeler için önerilen “Çok Boyutlu Uyarlanmış Testler-Multidimensional Adaptive Testing”, Çoklu test bataryaları ile BUT uygulamaları ve çok boyutlu MTK ile soru bankası geliştirmeye yönelik ileri çalışmaların yapılması önerilebilir.

KAYNAKÇA

- Aarts, R., Steidel, K., Manuel, B. A. F., Driessen, E. W. (2010). Progress Testing in Resource-Poor Countries: A Case from Mozambique. *Medical Teacher*, 32: 461-463.
- Ackerman, T. A. and Davey, T. C. (1991). Concurrent Adaptive Measurement of Multiple Abilities. The Annual Meeting of the American Educational Research Association'da sunulan bildiri, Chicago.
- Albano, M.G., Cavallo, F., Hoogenboom, R., Magni, F., Majoor, G., Manenti, F., ve diğeri. (1996). An International Comparison of Knowledge Levels of Medical Students: The Maastricht Progress Test. *Medical Education*, 30:239-245.
- Altman, D.G., Bland, J.M. (1983). Measurement in Medicine: The Analysis of Comparison Studies. *Statistician*. 32:307-317
- Al Alwan, I., Al-Moamary, M., Al-Attas, N., Al Kushi, A., ALBanyan, E., Zamakhshary, M., ve diğeri. (2011). The Progress Test as a Diagnostic Tool for a New PBL Curriculum. *Education for Health*. 24 (3): 493-498..
- Andrich, D., Hagquist, C. (2011). Real and Artificial Differential Item Functioning. *Journal of Educational and Behavioral Statistics*. 000 (00), pp. 1-30. <http://jebs.aera.net>
- Asril, A., Marais, I. (2011). Applying a Rasch Model Distractor Analysis Applications of Rasch. *Measurement in Learning Environments Research*. 2: 77-100
- Ateş, C. (2009). Sağlık Araştırmalarında Sınıf İçi Korelasyon Katsayısının Kullanımı. *Türkiye Klinikleri Journal of Biostatistic*. 1 (2):59-64
- Baykul, Y. (2000). *İstatistik Metodlar ve Uygulamalar*. Anı Yayıncılık: Ankara.
- Berkel, H.J.M., Nuy, H.J.P., Geerligs, T. (1995). The Influence of Progress Tests and Block Tests on Study Behaviour. *Instructional Science*. 22:317-322.
- Birenbaum, M., Dochy, F.J.R.C. (1996). *Alternatives in Assessment of Achievements, Learning Processes and Prior Knowledge*. Kluwer Academic Publishers. Boston, London.
- Bjorner, J.B., Chang, C.H., Thissen, D., Reeve, B.B. (2007). Developing Tailored Instruments: Item Banking and Computerized Adaptive Assessment. *Quality of Life Research*. 16 (Suppl). 95-108.

- 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. http://eric.ed.gov/ERICDocs/data/ericdocs2/content_storage_01/0000000b/80/0d/e1/36.pdf].
- Blake, J. M., Norman, G. R., Keane, D. R., Mueller, C. B., Cunningham, J., Didyk, N. (1996). Introducing Progress Testing in McMaster University's Problem-Based Medical Curriculum: Psychometric Properties and Effect on Learning. *Academic Medicine*. 71(9):1002-7.
- Bland, J.M., Altman, D.G. (1999). Measuring Agreement in Method Comparison Studies. *Statistical Methods in Medical Research*. 8:135–160.
- Bland, J.M., Altman, D.G. (1986). Statistical Methods For Assessing Agreement Between Two Methods of Clinical Measurement. *Lancet*. 1: 307-10
- Bridge, P.D., Musial, J., Frank, R., Roe, T., Sawilowsky, S. (2003). Measurement Practices: Methods for Developing Content-Valid Student Examinations. *Medical Teacher*. 25:414–421.
- Brown, J.D. (1997). Computers in Language Testing: Present Research and Some Future Directions. *Journal of Language Learning & Technology*. 1:44-49
- Boston, C. (2002). The Concept of Formative Assessment. *Practical Assessment, Research & Evaluation*, 8(9). <http://pareonline.net/getvn.asp?v=8&n=9>
- Boud, D. (2007). *Reframing Assessment as if Learning was Important*. in *Rethinking Assessment in Higher Education: Learning For The Longer Term*. D. Boud and N. Falchikov (Ed.). 14–25. London: Routledge.
- Bulut, O., Kan, A. (2012) Application Of Computerized Adaptive Testing to Entrance Examination for Graduate Studies in Turkey. *Eurasian Journal of Educational Research*. 49: 61-80
- Camilli, G., Shepard, L.A. (1994). *Methods for Identifying Biased Test Items*. Sage Publications, California.
- Chang, C.H. (2004). Developing Tailored Instruments: Item Banking and Computerized Adaptive Assessment. *Advances in Health Outcomes Measurement Conference*, Bethesda, MD sunulan bildirir. <http://www.outcomes.cancer.gov/conference/irt/chang.pdf>
- Chung, W.W. (2010). Recent Developments in Rasch Measurement. *The Hong Kong Institute of Education*. <http://www.ied.edu.hk/ps/view.php?secid=692>
- Coombes, L., Ricketts, C., Freeman, A., Stratford, J. (2010). Beyond Assessment: Feedback for Individuals and Institutions Based on the Progress Test. *Medical Teacher*. 32:486-490.

- Crane, P.K., Hart, D.L., Gibbons, L.E., Cook, K.F. (2006). 37-Item Shoulder Functional Status İtem Pool had Negligible Differential Item Functioning. *Journal of Clinical Epidemiology*.18:54:35
- Çıkrıkçı-Demirtaşlı, N. (1999). Psikometride Yeni Ufuklar: Bilgisayar Ortamında Bireye Uyarlanmış Test. *Türk Psikoloji Bülteni*. 5 (13): 31-36.
- De Champlain, A., Cuddy, M. M., Scoles, P. V., Brown, M., Swanson, D. B., Holtzman, K. (2010). Progress Testing in Clinical Science Education: Results of a Pilot Project Between the National Board of Medical Examiners and a US Medical School. *Medical Teacher*. 32:503-508.
- Dent, J., Harden, R. (2005). *Practical Guide for Medical Teacher*. Elsevier, Edinburg, London, Second Edition.
- Demirtaşlı, N. ve Arıkan, S. (2009). ÖİS’de MTK Uygulamaları. I. Ulusal Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi’inde sunulmuş bildiri. Ankara Üniversitesi Yayınları. 241: 225–227.
- Downing, S.M., Haladyna, T.M. (2004). Validity Threats: Overcoming Interference with Proposed Interpretations of Assessment Data. *Medical Education*. 38:327–333.
- Doğanay Erdoğan B. (2012). Çoklu Atama Yöntemlerinin Rasch Modelleri İçin Performansının Benzetim Çalışması İle İncelenmesi. Ankara Üniversitesi, Sağlık Bilimleri Enstitüsü, Doktora Tezi.
- Eggen, T.J.H.M. ve Straetmans, G.J.J.M. (2000). Computerized Adaptive Testing for Classifying Examinees into Three Categories. *Educational and Psychological Measurement*. 60(5): 713-734
- Elhan, A.H. (2002). *Rasch Analizinin İncelenmesi ve Fiziksel Tıp ve Rehabilitasyon Alanında Bir Uygulama*. (Doktora Tezi). Ankara Üniversitesi, Sağlık Bilimleri Enstitüsü.
- Embretson S.E., Reise S.P (2000). *Item Response Theory for Psychologists*. Lawrence Erlbaum Associates, Publishers
- Freeman A., van der Vleuten, Nouns, Z., Ricketts C., Welbourn, R.B. (2010). Progress Testing Internationally. *Medical Teacher*. 32:451- 455.
- Finucane, P., Flannery, D., Keane, D., Norman, G. (2010). Cross-institutional Progress Testing: Feasibility and Value to a New Medical School. *Medical Education*. 44:184-186.
- Georgiadou E., Triantafillous E., Economides A. A. (2006). Evaluation Parameters for Computer-Adaptive Testing. *British Journal of Educational Technology*. 37:2 61-278.

Greenstein, L. (2010). *What Teachers Really Need to Know About Formative Assessment*. ASCD Alexandria, Virginia USA.

GMAT Official Web Site (2013). <http://www.mba.com/mba/thegmat>.

GRE Web Sitesi (2013). Frequently Asked Questions. [Online] http://www.ets.org/gre/revised_general/faq/

Hambleton, R.K., Swaminathan, H., Rogers, H.J. (1991). *Fundamentals of Item Response Theory*. Sage Publications.

Hammerman, E. L. (2009) *Formative Assessment Strategies for Enhanced Learning in Science, K-8*. Corwin Press, USA.

Hays, R.D., Morales, L.S., Reise, S.P., (2000). Item Response Theory and Health Outcomes Measurement in the 21st Century. *Medical Care*. 38 (9 suppl.II): 28-42.

Heritage, M. (2007). Formative Assessment: What Do Teachers Need to Know and do? *Phi Delta Kappan*. 89 (2): 140–145.

Hagquist, C., (2010). Resolving Differential Item Functioning by Split of Items. ACSPRI Conferences, ACSPRI Social Science Methodology Conference'da sunulan bildiri. <http://conference.acspri.org.au/index.php/conf/2010/paper/view/115>

Hols-Elders, W., Bloemendaal P., Bos, N., Quaak, M., Sijtermans, R., De Jong, P. (2008). Twelve Tips for Computer-Based Assessment in Medical Education. *Medical Teacher*. 2008; 30: 673–678

International Foundations of Medicine (2011). <http://www.nbme.org/Schools/iFoM/index.html>

International Partnership for Progress Testing (2011). <http://ipptx.org/>

İşeri, A. I. (2002). *Assessment of Students' Mathematics Achievement Through Computer Adaptive Testing Procedures*. (Yayınlanmamış Doktora Tezi). Middle East Technical University, Turkey.

Jaradat, D. & Tollefson, N. (1988). The Impact of Alternative Scoring Procedures for Multiple-Choice Items on Test Reliability, Validity and Grading. *Educational and Psychological Measurement*. 48, 627-635

Kalender, İ. (2009). Başarı ve Yetenek Kestirimlerinde Yeni Bir Yaklaşım: Bilgisayar Ortamında Bireyselleştirilmiş Testler (Computerized Adaptive Tests - CAT). *Cito eğitim: kuram ve uygulama*. 5: 39-48

- Kaptan, F. (1993). Yetenek Kestiriminde Adaptive (Bireyselleştirilmiş) Test Uygulaması ile Geleneksel Kağıt-Kalem Testi Uygulamasının Karşılaştırılması. (Yayınlanmamış Doktora Tezi). Hacettepe University. Turkey.
- Karay. Y., Schaubert, S.K., Stosch, C., Schuettpelz-Brauns, K. (2012). Can Computer-Based Assessment Enhance The Acceptance of Formative Multiple Choice Exams? A utility Analysis. *Medical Teacher*. 34(4):292-6
- Kingsbury, G. and Hourser, R. (1993). Assessing the Utility of Item Response Models: Computer Adaptive Testing. *Journal of Educational Measurement*.12:21-27
- Kreiter, C.D., Ferguson, K., Gruppen, L. D. (1999), Evaluating the Usefulness of Computerized Adaptive Testing for Medical in-course Assessment. *Academic Medicine*. 74(10):1125-8
- Langer, M.M., Swanson, D.B. (2010). Practical Considerations in Equating Progress Tests. *Medical Teacher*. 32: 509–512
- Linacre, J.M. (2000). Computer-Adaptive Testing: A Methodology Whose Time Has Come. <http://www.rasch.org/memo69.pdf>.
- Lunz, M. E., Bergstrom, B. A. (1994). An Empirical Study of Computerized Adaptive Test Administration Conditions. *Journal of Educational Measurement*. 31(2): 251-263.
- McHarg J., Bradley P., Chamberlain S., Ricketts C., McLachlan J.C. (2005). Assessment of Progress Tests. *Medical Education*. 39: 221–227.
- Miller, V. D. (2003). *Assessment of Student Achievement: a Comparative Study o Student Achievement Using Paper and Pencil Assessment and Computerize Dadaptive Testing*. (Yayınlanmamış Doktora Tezi). Wayne State University.Graduate School.
- Muijtjens, A.M.M, van Mameren, H., Hoogenboom, R.J.I., Evers, J.L.H., van der Vleuten, C.P.M. (1999). The Effect of a 'Don't Know' Option on Test Scores: Number-Right and Formula Scoring Compared. *Medical Education*. 33:267–275.
- Muijtjens, A.M., Schuwirth, L.W., Cohen-Schotanus, J., Thoben, A.J., van der Vleuten, C.P. (2008). Benchmarking by Cross-Institutional Comparison of Student Achievement in a Progress Test. *Medical Education*. 42(1):82–88.
- Muijtjens, A.M.M., Wijnen W. (2010). *Progress Testing*. (eds) Van Berkel H, Scherpbier A, Hillen H, Van Der Vleuten C, Lessons from Problem-Based Learning. Oxford: Oxford University Press.

- Muijtjens, A.M.M, Hoogenboom R.J.I, Verwijnen, G.M., van der Vleuten, C.P.M. (1998). Relative or Absolute Standards in Assessing Medical Knowledge using Progress Test. *Advances in Health Science Education*. 3:81–87.
- Molina, G.J. (2008). Modeling Item Banking: Analysis and Design of a Computerized System. *Revista Electrónica de Metodología Aplicada*. 3(2): 1-14
- Norman, G., Neville, A., Blake, J., Mueller, B. (2010). Assessment Steers Learning Down the right road: Impact of Progress Testing on Licensing Examination Performance. *Medical Teacher*. 32:496–499.
- Nouns, Z. M., & Georg, W. (2010). Progress Testing in German Speaking Countries. *Medical Teacher*. 32:467-470.
- Öztuna, D. (2008). *Kas İskelet Sistemi Sorunlarının Özürlülük Değerlendirmesinde Bilgisayar Uyarlamalı Test Yönteminin Uygulanması*. (Doktora Tezi). Ankara Üniversitesi, Sağlık Bilimleri Enstitüsü.
- 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: 1–18.
- Rademakers, J., Cate, T.J., Bär, P.R. (2005). Progress Testing With Short Answer Questions. *Medical Teacher*. 27(7):578-82.
- Raju, N. S. (1988). The Area Between Two Item Characteristic Curves. *Psychometrika* 53, 495–502
- Reckase, M. D. (2003). Item Pool Design for Computerized Adaptive Tests. Annual meeting of the National Council on Measurement in Education'da sunulmuş bildirisi. Chicago, IL.
- Ricketts, C., Freeman, A., Pagliuca, G., Coombes, L., Archer, J., (2010). Difficult Decisions for Progress Testing: How Much and How Often? *Medical Teacher*. 32: 513–515.
- Rodney, G. L., ve Drasgow, F. (1990). Evaluation of Two Methods for Estimating Item Response Theory Parameters When Assessing Differential Item Functioning. *Journal of Applied Psychology*. 75, 164- 174.
- Roediger, H.L., Karpicke, J.D. (2006). Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention. *Psychological Science*. 17:249–255.
- Roediger, H.L., Butler, A.C. (2010.) The Critical Role of Retrieval practice in Longterm Retention. *Trends Cognitive Science*. 15:20–27.

- Roex, A., Degryse J. (2004). A Computerized Adaptive Knowledge Test as an Assessment Tool in General Practice: a Pilot Study. *Medical Teacher*, 26 (2): 178–183
- Rudner, L.M. (1998). An On-Line, Interactive, Computer Adaptive Testing Tutorial. <http://edres.org/scripts/cat>.
- RUMM LABORATORY PTY LTD. (2007). RUMM2020, Version 4.1. Duncraig, WA: RUMM Laboratory Pty Ltd.
- Sands, W. A., Waters, B. K., & McBride, J. R. (Eds.) (1997). *Computerized Adaptive Testing: from Inquiry to Operation*. Washington. DC: American Psychological Association.
- Schaap, L., Schmidt, H., & Verkoeijen, P. J. L. (2012). Assessing Knowledge Growth in a Psychology Curriculum: Which Students Improve Most? *Assessment & Evaluation in Higher Education*. 37(7): 875-887
- Schauber, S., Nouns, Z.B. (2010). Using the Cumulative Deviation Method for Cross-institutional benchmarking in the Berlin Progress Test. *Medical Teacher*. 32:471–475.
- Schuwirth, L., Bosman, G., Henning, R. H., Rinkel, R., & Wenink, A. C. G. (2010). Collaboration on Progress Testing In Medical Schools in the Netherlands. *Medical Teacher*. 32: 476-479.
- Sick, J. (2008). Rasch Measurement in Language Education. 12 (1): 1-6. http://jalt.org/test/sic_1.htm.
- Sijtsma K., van der Ark, L.A.. (2003). Investigation and Treatment of Missing Item Scores in Test and Questionnaire Data. *Multivariate Behavioral Research* 38(4):505-528.
- Simms, L. J., & Clark, L. A. (2005). Validation of a Computerized Adaptive Version of the Schedule for Nonadaptive and Adaptive Personality. *Psychological Assessment*. 17:28-43.
- Swaminathan, H., Rogers, H. J. (1990). Detecting Differential Item Functioning Using Logistic Regression Procedures. *Journal of Educational Measurement*. 27: 361-370.
- Swanson, D. B., Holtzman, K. Z., Butler, A., Langer, M. M., Nelson, M. V., Chow, J. W. M., et al. (2010). Collaboration Across The Pond: The Multi-School Progress Testing Project. *Medical Teacher*. 32: 480-485.
- Şencan, H. (2005). *Sosyal ve Davranışsal Ölçümlerde Güvenilirlik ve Geçerlilik*. Ankara: Seçkin Yayıncılık.

- Tavakol, M., Dennick, R. (2013). Psychometric Evaluation of a Knowledge Based Examination Using Rasch Analysis: An illustrative Guide: AMEE Guide No. 72. *Medical Teacher*. 35: e838–e848.
- Tekin, H. (2009). *Eğitimde ölçme ve değerlendirme*. Yargı Yayınları. Ankara.
- van Ginkel, J. R., & Van der Ark, L. A. (2005). SPSS syntax for Missing Value Imputation in Test and Questionnaire Data. *Applied Psychological Measurement*. 29:152-153.
- Thissen, D., Reeve, B.B., Bjorner, J.B., & Chang, C.-H. (2007). Methodological Issues for Building Item Banks And Computerized Adaptive Scales. *Quality of Life Research*. 16:109-116.
- TOEFL Web Sitesi, 2013. FAQs About the Computer-Based TOEFL. Test[Online]. <http://www.ets.org/toefl/ibt/faq>.
- van der Vleuten, C.P.M., Verwijnen, G.M., Wijnen, W.H.FW. (1996 a). Fifteen Years of Experience with Progress Testing in a Problem-Based Learning Curriculum. *Medical Teacher*. 18(2):103–110.
- van der Vleuten, C.P.M. (1996b). The Assessment of Professional Competence: Developments, Research and Practical Implications. *Advance in Health Science Education*. 1:41–67.
- van Diest, R., Van Dalen, J., Bak M, Schruers, K., van der Vleuten, C., Muijtjens, A., Scherpbier A. (2004). Growth of Knowledge in Psychiatry and Behavioural Sciences in a Problem-Based Learning curriculum. *Medical Education*. 38(12):1295–1301.
- van der Linden, W. J., & Glas, C. A. W. (2000). *Computerized Adaptive Testing: Theory and Practice*. Dordrecht, the Netherlands: Kluwer.
- van der Vleuten, C.P.M., Schuwirth, L.W., Muijtjens, A.M., Thoben, A.J., Cohen-Schotanus, J., Van Boven, C.P. (2004). Cross Institutional Collaboration in Assessment: A Case on Progress Testing. *Medical Teacher* 26(8):719–725.
- Verhoeven, B.H., Verwijnen G.M., Scherpbier A.J., van der Vleuten C.P.M. (2002). Growth of Medical Knowledge. *Medical Education*. 36(8):711–717.
- Verhoeven, B. H., Snellen-Balendong, H. A. M., Hay, I. T., Boon, J. M., Van Der Linde, M. J., Blitz-Lindeque, J. J., ve diğerleri. (2005). The Versatility of Progress Testing Assessed in an International Context: a Start for Benchmarking Global Standardization? *Medical Teacher*. 27(6), 514-520.
- Wade, L., Harrison, C., Hollands, J., Mattick, K., Ricketts, C., Wass, V. (2012). Student Perceptions of the Progress Test in Two Settings and the

Implications For Test Deployment. *Advances in Health Science Education*. 17:573–583

- Wainer, H., Dorans, N.J., Flaugher, 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. *Archives of Physical Medicine and Rehabilitation*. 84 (Suppl. 2): S43–51.
- Wang T., Vispoel W.P. (1998). Properties of Ability Estimation Methods in Computerized Adaptive Testing. *Journal of Educational Measurement*. 35 (2):, 109–135
- Wang T., Kolen M.J. (2001). Evaluating Comparability in Computerized Adaptive Testing: Issues, Criteria and an Example. *Journal of Educational Measurement*. 38 (1): 19-49
- Weiss, D.J. (2011). Better Data From Better Measurements Using Computerized Adaptive Testing. *Journal of Methods and Measurement in the Social Sciences*, 2 (1), 1-27.
- Weiss, D. J., Kingsbury, G. G. (1984). Application of Computerized Adaptive Testing to Educational Problems. *Journal of Educational Measurement*. 21: 361-375.
- Weiss, D.J. (2004). Computerized Adaptive Testing for Effective and Efficient Measurement in Counseling and Education. *Measurement and Evaluation in Counseling and Development*. 37 (2):70-84
- Weiss, D. J. (2010). CAT Central: A global resource for Computerized Adaptive Testing Research and Applications. <http://www.psych.umn.edu/psylabs/CATCentral>.
- Wrigley, W., van der Vleuten., C.P.M., Freeman, A., Muijtjens, A. (2012). A Systemic Framework for the Progress Test: Strengths, Constraints and Issues: AMEE Guide No. 71. *Medical Teacher*. 34: 683–697
- Zimmerman, D.W., Williams R.H. (2003). A New Look at the Influence of Guessing on the Reliability of Multiple-Choice Tests. *Applied Psychological Measurement*. 27:357–371
- Zitny, P. (2011). Computerized Adaptive Testing. Precision, Validty and Efficiency). *Ceskoslovenska Psychologie*. 55(2): 167-179
- Zitny ,P., Halama, P., Jelinek, M., Kveton, P. (2012). Validty of Cognitive Ability Test Comparison of Computerized Adaptive Testing with Paper and Pencil And

Computer-Based Forms of Administration. *Studia Psychologica*. 54: 181-194

Zumbo, B. D. (1999). *A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores*. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense. <http://www.educ.ubc.ca/faculty/zumbo/DIF/handbook.pdf>

EKLER

EK 1.

Çizelge 6. Gelişim Sınavı Uygulamaları Kullanılma Durumu

Okul	Tanım
Hollanda grubu; Hollanda da 5 Tıp Fakültesi (Groningen, Leiden, Maastricht, Nijmegen ve VU Amsterdam) Belçika'da Ghent Üniversitesi	1977'de Maastricht'te başlamış ve 1999 yılında adı geçen diğer tıp fakülteleri ile işbirliğine başlanmıştır. Sınav biçimlendirmeye ve belgelendirmeye yönelik olarak uygulanmaktadır. <i>Kağıt kalem testi</i> olarak uygulanmaktadır. Tek doğru cevaplı 200 çoktan seçmeli soru ile yapılmaktadır (Şans başarısı için düzeltme formülü kullanılmaktadır). Sınav dört saat sürmektedir. Yılda dört kez uygulanmaktadır. Sorular tüm tıp eğitimi bilgi kapsamını içeren bir belirtke tablosu ile belirlenmektedir. 5000 sorudan oluşan bir soru havuzu mevcuttur.
Kanada, McMaster (Mezuniyet öncesi tıp eğitimi ve mezuniyet sonrası tıp eğitimini kapsayacak şekilde), Limerick Üniversitesi	1992 yılında uygulanmaya başlanmıştır. Sınav biçimlendirmeye ve belgelendirmeye yönelik olarak uygulanmaktadır. Sınav çevrimiçi olarak uygulanmaktadır. Tek doğru cevaplı 180 çoktan seçmeli soru ile yapılmaktadır (Şans başarısı için düzeltme formülü kullanılmaktadır). 540 öğrenciye uygulanmaktadır (McMaster). Sınav üç saat sürmektedir. Yılda üç kez uygulanmaktadır. 3000 maddeden oluşan bir soru havuzu mevcuttur.

Çizelge 6. Gelişim Sınavı Uygulamaları Kullanılma Durumu (Devamı)

Okul	Tanım
Almanya- Berlin grubu; Almanya – Berlin Regel, Witten, Aachen, Bochum, LMU Munich, Köln, Munster, Hannover, Mannheim, Regensburg, Austria – Graz, Innsbruck	1999 yılından bu yana uygulanmaktadır. Sınav biçimlendirmeye yönelik olarak uygulanmaktadır. <i>Kağıt kalem testi</i> olarak uygulanmaktadır. Tek doğru cevaplı 200 çoktan seçmeli soru ile yapılmaktadır (Şans başarısı için düzeltme formülü kullanılmaktadır). Sınav üç saat sürmektedir. Yılda iki kez uygulanmaktadır. Her uygulamada yaklaşık 8500 öğrenci sınava girmektedir. 5300 maddeden oluşan bir soru havuzu mevcuttur.
Ulusal Tıp Sınavları Kurulu-1 (National Board of Medical Examiners-NBME 1) Barts, St. George's London, Leeds ve Queens Üniversitesi, Belfast	Sınavlar İngiltere'deki dört okul ve Ulusal Tıp Sınavları Kurulu-NBME (ABD) ortaklığı ile 2008 yılından bu yana yapılmaktadır. Her okul sınavı farklı amaçlar için (belgelendirmeye, biçimlendirmeye yönelik) uygulamaktadır. NBME web üzerinden uygulamaktadır. Tek doğru cevaplı 200 çoktan seçmeli soru ile yapılmaktadır (Şans başarısı için düzeltme formülü kullanılmaktadır). Sınav genellikle beş yıllık eğitim döneminin son üç yılındaki öğrencilerine uygulanmaktadır. Sınav üç saat sürmektedir. Yılda iki kez uygulanmaktadır.
Ulusal Tıp Sınavları Kurulu-2 (National Board of Medical Examiners-NBME 2) South Florida Üniversitesi ve Case Western Reserve Üniversitesi	ABD'deki iki okulun işbirliği ile yapılmaktadır. South Florida Tıp Fakültesinde sınav biçimlendirmeye yönelik olarak uygulanmaktadır. Sınav web üzerinden uygulamaktadır. Tek doğru cevaplı 230 çoktan seçmeli soru ile yapılmaktadır Sınav beş saat sürmektedir. Yılda dört kere uygulanmaktadır.

Çizelge 6. Gelişim Sınavı Uygulamaları Kullanılma Durumu (Devamı)

Okul	Tanım
Southern Illinois Üniversitesi, New Mexico Vanderbilt, Üniversitesi, Penn State, Texas Tech, Georgia Tıp Okulu, Minnesota Üniversitesi	<p>2004 yılından bu yana bu altı tıp fakültesinin işbirliği ile tek bir sınav olarak yapılmaktadır.</p> <p>Sınav biçimlendirmeye yönelik olarak uygulanmaktadır.</p> <p><i>2010 yılına kadar kağıt kalem testi</i> olarak uygulanmış (çoktan seçmeli 70 soru ile), 2010 yılından bu yana çevrimiçi olarak uygulanmaktadır.</p> <p>Tek doğru cevaplı 230 çoktan seçmeli soru ile yapılmaktadır (Şans başarısı için düzeltme yapılmamaktadır).</p> <p>Dört yıllık eğitim dönemindeki 3200 öğrenci katılmaktadır. Birkaç yüz maddeden oluşan bir soru havuzu vardır.</p>
Manchester, İngiltere	<p>Manchester Üniversitesinde gelişim sınavı 1997 yılından bu yana uygulanmaktadır.</p> <p>Sınav belgelendirmeye yönelik olarak uygulanmaktadır.</p> <p><i>Kağıt kalem testi</i> olarak uygulanmaktadır.</p> <p>Tek doğru cevaplı 125 çoktan seçmeli madde ile yapılmaktadır (Şans başarısı için düzeltme yapılmamaktadır).</p> <p>Sınav iki saat sürmektedir.</p> <p>Yılda iki kere uygulanmaktadır.</p> <p>Beş yıllık eğitim döneminde bulunan 2000 öğrenciye uygulanmaktadır.</p> <p>Maddeler İngiltere'deki 14 tıp fakültesinin Konsorsiyumu ile oluşturulan, Universities Medical Assessment Partnership-UMAP soru bankasından elde edilmektedir.</p>

Çizelge 6. Gelişim Sınavı Uygulamaları Kullanılma Durumu (Devamı)

Okul	Tanım
Peninsula Tıp Fakültesi ve Diş Hekimliği Fakültesi, İngiltere	<p>Sınav 2002 yılından bu yana uygulanmaktadır.</p> <p>Sınav belgelendirmeye yönelik olarak uygulanmaktadır.</p> <p>Başlangıçta kağıt kalem testi olarak uygulanmakta olan sınav 2009'dan bu yana çevrimiçi olarak uygulanmaktadır.</p> <p>Tek doğru cevaplı 125 çoktan seçmeli madde ile yapılmaktadır (Şans başarısı için düzeltme yapılmaktadır).</p> <p>Sınav üç saat sürmektedir.</p> <p>Yılda iki kez uygulanmaktadır.</p> <p>Beş yıllık eğitim dönemindeki 900 öğrenci katılmaktadır.</p> <p>2500 maddeden oluşan bir soru havuzunun yanı sıra UMAP soru havuzu da kullanılmaktadır.</p> <p>Maddeler tüm tıbbi bilgi kapsamını içeren bir belirtke tablosu ile belirlenmektedir.</p>
Tampere, Finlandiya	<p>Sınav biçimlendirmeye yönelik olarak uygulanmaktadır.</p> <p>Kağıt kalem testi olarak uygulanmaktadır.</p> <p>Doğru/Yanlış formatında 224 madde ile yapılmaktadır (Şans başarısı için düzeltme yapılmamaktadır). Sınav üç saat sürmektedir. Tüm eğitim dönemlerinden 600 öğrenci katılmaktadır. 10000 maddeden oluşan bir soru havuzunun vardır. Maddeler tüm tıbbi bilgi kapsamını içeren bir belirtke tablosu ile belirlenmektedir.</p>
Otago Tıp Fakültesi, Yeni Zellanda	<p>Sınav biçimlendirmeye yönelik olarak uygulanmaktadır.</p> <p>Çevrimiçi olarak uygulanmaktadır.</p> <p>Tek doğru cevaplı 100 çoktan seçmeli soru ile yapılmaktadır.</p> <p>Yılda iki kez uygulanmaktadır</p>
Sao Paulo Tıp Fakültesi, Brezilya	<p>2001 yılından bu yana uygulanmaya başlanmıştır.</p> <p>Tek doğru cevaplı 100 çoktan seçmeli soru ile yapılmaktadır (Şans başarısı için düzeltme yapılmamaktadır).</p> <p>2000 maddeden oluşan bir soru havuzu mevcuttur.</p>

Çizelge 6. Gelişim Sınavı Uygulamaları Kullanılma Durumu (Devamı)

Okul	Tanım
Endonezya Üniversitesi Tıp Fakültesi	<p>2008 yılında uygulanmaya başlanmıştır.</p> <p>Sınav biçimlendirmeye yönelik olarak uygulanmaktadır.</p> <p><i>Kağıt kalem testi</i> olarak uygulanmaktadır.</p> <p>Tek doğru cevaplı 120 çoktan seçmeli soru ile yapılmaktadır (Şans başarısı için düzeltme formülü kullanılmaktadır).</p> <p>1200 öğrenciye uygulanmaktadır.Yılda iki kez uygulanmaktadır.1200 maddeden oluşan soru havuzu vardır.</p>
Mozambique Katolik Üniversitesi	<p>2001 yılında uygulanmaya başlanmıştır.</p> <p>Sınav biçimlendirmeye ve belgelendirmeye yönelik olarak uygulanmaktadır. <i>Kağıt kalem testi</i> olarak uygulanmaktadır.</p> <p>Tek doğru cevaplı 200 çoktan seçmeli madde ile yapılmaktadır (Şans başarısı için düzeltme formülü kullanılmaktadır).</p> <p>200 öğrenciye uygulanmaktadır. Yılda dört kez uygulanmaktadır. Sınav dört saat sürmektedir.</p>

Not: Sınavı belgelendirmeye yönelik değerlendirme amacı ile yaptığı belirtilen fakülteleri aynı zamanda sınavı sonuçlarını biçimlendirmeye yönelik değerlendirmeye olarak kullanmaktadırlar. İlk değerlendirme için kullanılmamaktadır (Freeman A. ve diğerleri, 2010).

EK.2

Çizelge 7. MFF Gösteren Maddelerin Dağılımı

Madde	Anabilim Dalı	Dönem
s2- A. radialisten el bileğinde nabız alabilmek için aşağıda eşleştirilen hangi iki kas tendonu arasından palpasyon yapılmamalıdır?(2)	ANATOMİ	2
s4- N. musculocutaneous hasarı olan bir hastada aşağıdaki ön kola en kuvvetli fleksiyon yaptıran kas aşağıdakilerden hangisidir?(5)	ANATOMİ	1 (-5,-4)
s6- Radikal prostatektomi geçiren bir hastada m. sphincter urethra externus'un paralizisi sonucunda hastada üriner inkontinans meydana gelmiştir. Cerrahi sırasında zarar görmüş olan sinir aşağıdakilerden hangisidir? (8)	ANATOMİ	2 (-5,-1)
s8- Humerus gövde kırıklarından sonra hastada aşağıdaki hareket problemlerinden hangisiyle karşılaşma olasılığımız daha yüksektir?(10)	ANATOMİ	1
s9- Asansörde ayakta duran bir kişide asansörün ivmeli hareket yaptığına ilişkin enformasyon aşağıdaki bölgelerden hangisindeki reseptörlerin uyarılmasına bağlıdır? (12)	FİZYOLOJİ	2 (-3,-4)
s10- Puberte sırasında hipotalamus-hipofiz-gonad aksının matür hale gelmesi ile ilgili olarak hangisi doğrudur? (13)	FİZYOLOJİ	2
s11- Akut bir kanamaya yanıt olarak baroreseptör refleksi tetiklendiğinde, aşağıdakilerden hangisinde kanama öncesindeki değerinin üzerinde bir artış beklenir? (15)	FİZYOLOJİ	-1
s12- Aşağıdakilerden hangisi plazma PO_4^{3-} konsantrasyonu azaldığında devreye giren regülasyon mekanizmalarından biri değildir ? (19)	HİSTOLOJİ	2
s13- Bir hormonun hücrelerarası maddeye salgılanarak yakın çevrede gösterdiği etki nasıl isimlendirilir? (21)	HİSTOLOJİ	2 (-1)
s15- Kan beyin bariyerinin yapısal temelini hangisi oluşturur? (24)	HİSTOLOJİ	1,3 (-5,-4)
s18- Bağ dokusu ile ilgili açıklamalardan doğru olanı işaretleyiniz.(30)	HİSTOLOJİ	1 (-5,-4)
s19- EKG için aşağıdaki ifadelerden hangileri doğrudur? (31)	BİYOFİZİK	2 (-4,-1)
s21- Mikrosatellit instabilitesi testi aşağıdaki gastrointestinal kanserlerden hangisi için bir tarama testidir? (35)	GENETİK	3 (-5)
s25- Distrofin proteininin fonksiyonu aşağıdakilerden hangisidir? (41)	GENETİK	3 (-5,-4,-1)
s26- Hemoglobinin oksijene affinitesini azaltan 2,3 bisfosfoglisarat hangi metabolik yolun bir ara ürünüdür? (43)	BİYOKİMYA	-4, -5
s27- Sindirim sırasında bağırsaktan emilen uzun zincirli yağ asitleri (12 karbondan uzun) bağırsak hücresinin içinde nasıl bir işlemde geçerler? (44)	BİYOKİMYA	1 (-5,-4)
s28- Aşağıdakilerden hangisi enzimatik bir reaksiyonun hızını etkileyen faktör değildir ? (45)	BİYOKİMYA	1 (-5,-4)
s30- Aşağıdakilerin hangisi akut miyokard enfarktüsünün tanısında yararlanılan laboratuvar testleri arasında yer almaz ? (48)	BİYOKİMYA	3
s35- Trafik kazası nedeniyle 10 gündür Yoğun Bakım Ünitesi'nde yatmakta olan ventilatöre bağlı bir hastada pnömoni gelişmiş, endotrakeal aspirat kültüründe kanlı agar besiyerinde beta hemolitik, yeşil metalik, parlak S tipi koloniler görülmüş ve besiyerinin üzüm gibi koktuğu dikkati çekmiştir. Yapılan incelemelerde bakterinin non-fermentatif, oksidaz pozitif ve gram negatif basil olduğu saptanmıştır.	MİKROBİYOLOJİ	
Bu olguda en olası mikroorganizma aşağıdakilerden hangisidir (56)		3,5 (-1)
s36- Aşağıdaki enterit etkeni bakterilerden hangisi yalnızca insanlarda infeksiyon oluşturur? (58)	MİKROBİYOLOJİ	3 (-5,-4)

Çizelge 7. MFF Gösteren Maddelerin Dağılımı (Devamı)

Madde	Anabilim Dalı	Dönem
s37- Bir gündüz bakımevinde, yaşları 3 ile 6 arasında değişen 80 çocuk ve 9 çalışana öğle yemeği için yemek firmasından getirilen kızarmış tavuklu pilav getirilmiştir. Yaklaşık 2 saat sonra yemeği yiyen 70 çocuğun 15'inde kusma, karın ağrısı, ishal şikayeti görülmüştür. Salgın sırasında kusma belirtisinin daha fazla olduğu gözlenmiş hastaların şikayetleri 6 saat içinde geçmiştir. Yemeğin mikrobiyolojik analizinde büyük, gri renkli, hemolizli, "R" koloni yapan sporlu basiller görülmüştür. Bu salgından sorumlu muhtemel mikroorganizma size göre aşağıdakilerden hangisidir? (59)	MİKROBİYO LOJİ	3
s38- Aşağıdaki virüslerden hangisi gözde hastalık etkeni olma olasılığı en düşüktür? (60)	MİKROBİYO LOJİ	3,5
s39- Aşağıdaki mikroorganizmalardan hangisi üretra normal florasında bulunmaz? (61)	MİKROBİYO LOJİ	2 (-1)
s40- Enterobakteri cinslerinden hangisi, güçlü üreaz aktivitesi nedeniyle üriner sistemde taş oluşturma ve kronikleşen üriner sistem infeksiyonları yapma eğilimi gösteren cinstir? (62)	MİKROBİYO LOJİ	2
s43- Lal kırmızı renginde ölü lekesi görüldüğünde cesette hangi zehirlenme için inceleme istenir? (69)	ADLİ TIP	4
s44- Fiilin anlam ve sonuçlarını kavrama kavramı ile ilgili tanımlamalardan hangisi öncelikle dikkate alınır. (70)	ADLİ TIP	4
s48- Bir fabrikada yönetici olarak çalışan 64 yaşındaki Ahmet Bey akut miyokard infarktüsü geçirdiği için hastanede yatmaktadır. Monitörden yakın izlemde olan hastanın PR aralıklarının normalden uzun, QRS komplekslerinin çoğunlukla normal konfigürasyonda olduğu ama zaman zaman QRS komplekslerinin P dalgalarını takip etmediği dikkati çekmektedir. Hastanede yatarken Ahmet Bey iki kez baygınlık da geçirmiştir. Ahmet Bey'in durumu aşağıdakilerden hangisi ile açıklanabilir? (87)	FİZYOPATO LOJİ	4,5 (-3)
s52- Myokard infarktüsü geçirmiş bir hastada, aşağıdaki antianjinal ilaçlardan hangisi göğüs ağrısının kronik profilaksisi için altın standart bir öneme sahiptir? (102)	FARMAKOL OJİ	4
s54- Atrial fibrilasyon için hangisi doğrudur? (104)	KARDİYOL OJİ	4,5
s55- Kalp damar hastalığı bulunan ve LDL kolesterol düzeyi uygun yaşam biçimi değişikliğine rağmen 160mm/dL olan bir hastada kolesterol düşürücü tedavi için birinci seçenek olarak kullanılması gereken ilaç grubu hangisidir? (105)	KARDİYOL OJİ	4,5
s57- Aşağıdakilerden hangisi erişkin hastalarda en sık primer peritonite yol açan bakteridir? (113)	GENEL CERRAHİ	5 1?
s59- 50 yaşın altındaki erişkinlerde kolon kökenli alt GİS kanamalarının en sık rastlanan nedeni aşağıdakilerden hangisidir? (115)	GENEL CERRAHİ	4
s64- Trafik kazası sonrası olay yerinde gördüğünüz bir yaralıda ciddi solunum sıkıntısı ile birlikte sağ hemitoraksın solunuma katılmadığını, boyunda venöz dolgunluk olduğunu ve boyunda palpasyonla trakeada sola deviasyon olduğunu saptadınız. Olası tanınız nedir? (122)	GÖĞÜS CERR.	5
s65- 24 yaşında kadın hastanın gebelik takibi sırasında yapılan rutin muayenesinde karaciğer enzimleri normal, HBs Ag pozitif ve HBV DNA negatif saptanmış. HBV infeksiyonunun vertikal geçiş riskini etkin bir şekilde azaltmak aşağıdakilerden hangisi önerilmelidir? (123)	GASTRO	4,5
s66-30 yaşındaki kadın hastanın üç yıldan beri önce sıvılara karşı, son dönemlerde ise katı ve sıvılara karşı gelişen yutma güçlüğü yakınması mevcut. Manometrik incelemesinde ise özofagus gövdesinde aperistaltizm, alt özofagus sfinkter basıncında yükseklik ve yetersiz gevşeme rapor edilmiş. Bu hastada tanınız nedir? (124)	GASTRO	4,5

Çizelge 7. MFF Gösteren Maddelerin Dağılımı (Devamı)

Madde	Anabilim Dalı	Dönem
s71-3 yaşında kız hastada, 39° C Ateş, serum C-reaktif protein yüksekliği ve tam idrar tahlilinde de; hematüri, bol lokosit, nitrit pozitifliği ve bakteri mevcut. Bu hastanın tanısı ve daha sonraki tanısal işlemleri ile ilgili olan en doğru ifade hangisidir? (131)	ÜROLOJİ	2,5
s78- Aşağıdakilerden hangisi alkol bağımlılığı tanısını koymada kullanılan ölçütlerden (fenomenlerden) değildir ? (148)	PSİKİYATRİ	4
s83- Aşağıdaki patolojilerin hangisi doğum eyleminin indüklenmesinin kesin kontrendikasyonudur (158)	KADIN DOĞUM	5
s87- Aşağıdaki yenidoğan konjunktivitlerinden hangisi 2–5. günlerde ortaya çıkar ve “oftalmia neonatoruma neden olur? (166)	ÇOCUK	5
s91 -Trafik kazası sonrası sağ bacakta deformite, kısalık, ağrı ve krepitasyon ile gelen hasta acil girişimleri tamamlanıp uzun bacak atel ile immobilize edildiği halde giderek artan bacak ağrısından şikayetçi. Bacakta aşırı şişlik mevcut. Nörovasküler muayene intakt. Pasif germe ile ağrıda artış mevcut. Tanınız nedir? (178)	ORTOPEDİ	5
s93- Aşağıdaki hangi yasaklı etken madde halter, güreş, boks gibi siklet sporcuları tarafından kısa süre içinde ağırlık düşmek amacıyla kullanılır? (181)	SPOR HEKİMLİĞİ	5
s97- Multisistem travmalı hastalarda <u>en sık</u> görülen şok tipi aşağıdakilerden hangisidir? (188)	ACİL	5
s103- Aşağıdakilerden hangisinde kronik oral ülserasyon görülür? (200)	DERMATOLOJİ	5

EK.3**Çizelge 8.** 103 Maddelik Soru Bankasında Yer Alan Maddelerin Sistem ve Bilim Alanlarına Göre Dağılımı

Madde No	Bilim alanı	Sistem
1 (m1)	ANATOMİ	Kan, kan yapıcı organlar, immun sistem
2 (m2)	ANATOMİ	Kas iskelet sistemi
3 (m3)	ANATOMİ	Dolaşım sistemi
4 (m5)	ANATOMİ	Kas iskelet
5 (m7)	ANATOMİ	Sindirim sistemi
6 (m8)	ANATOMİ	Ürogenital sistem
7 (m9)	ANATOMİ	Sinir sistemi
8 (m10)	ANATOMİ	Sinir sistemi
9 (m12)	FİZYOLOJİ	Sinir sistemi
10 (m13)	FİZYOLOJİ	Üriner sistem/Genital sistem
11 (m15)	FİZYOLOJİ	Dolaşım sistemi
12 (m19)	FİZYOLOJİ	Endokrin, metabolizma, beslenme
13 (m21)	HİSTOLOJİ	Endokrin, metabolizma, beslenme
14 (m22)	HİSTOLOJİ	Ürogenital sistem
15 (m24)	HİSTOLOJİ	Sinir sistemi
16 (m25)	HİSTOLOJİ	Dolaşım sistemi
17 (m26)	HİSTOLOJİ	Solunum sistemi
18 (m30)	HİSTOLOJİ	Deri bağ Dokusu
19 (m31)	BİYOFİZİK	Dolaşım sistemi
20 (m33)	BİYOFİZİK	Üriner sistem/Genital sistem
21 (m35)	GENETİK	Sindirim sistemi
22 (m36)	GENETİK	Ürogenital sistem
23 (m37)	GENETİK	Endokrin, metabolizma, beslenme
24 (m39)	GENETİK	Dolaşım sistemi
25 (m41)	GENETİK	Kas iskelet sistemi
26 (m43)	BİYOKİMYA	Kan, kan yapıcı organlar, immun sistem
27 (m44)	BİYOKİMYA	Sindirim sistemi
28 (m45)	BİYOKİMYA	Sinir sistemi
29 (m46)	BİYOKİMYA	Endokrin
30 (m48)	BİYOKİMYA	Dolaşım sistemi
31 (m50)	BİYOKİMYA	Solunum sistemi
32 (m52)	BİYOKİMYA	Kas iskelet sistemi
33 (m54)	MİKROBİYOLOJİ	Dolaşım
34 (m55)	MİKROBİYOLOJİ	Solunum sistemi
35 (m56)	MİKROBİYOLOJİ	Solunum
36 (m58)	MİKROBİYOLOJİ	Sindirim sistemi
37 (m59)	MİKROBİYOLOJİ	Endokrin, metabolizma, beslenme
38 (m60)	MİKROBİYOLOJİ	Duyu (Göz / KBB)
39 (m61)	MİKROBİYOLOJİ	Ürogenital sistem

Çizelge 8. 103 Maddelik Soru Bankasında Yer Alan Maddelerin Sistem ve Bilim Alanlarına Göre Dağılımı (Devamı)

Madde No	Bilim alanı	Sistem
40 (m62)	MİKROBİYOLOJİ	Ürogenital sistem
41 (m63)	ETİK	
42 (m65)	ETİK	
43 (m69)	ADLİ TIP	
44 (m70)	ADLİ TIP	
45 (m73)	BİYOİSTATİSTİK	
46 (m76)	HALK SAĞLIĞI	
47 (m79)	HALK SAĞLIĞI	
48 (m87)	FİZYOPATOLOJİ	Dolaşım sistemi
49 (m88)	FİZYOPATOLOJİ	Dolaşım sistemi
50 (m89)	FİZYOPATOLOJİ	Ürogenital sistem
51 (m92)	PATOLOJİ	Solunum sistemi
52 (m102)	FARMAKOLOJİ	Dolaşım sistemi
53 (m103)	FARMAKOLOJİ	Solunum
54 (m104)	KARDİYOLOJİ	Dolaşım sistemi
55 (m105)	KARDİYOLOJİ	Dolaşım sistemi
56 (m106)	KARDİYOLOJİ	Dolaşım sistemi
57 (m113)	GENEL CERRAHİ	Sindirim sistemi
58 (m114)	GENEL CERRAHİ	Sindirim sistemi
59 (m115)	GENEL CERRAHİ	Sindirim sistemi
60 (m116)	GENEL CERRAHİ	Sindirim sistemi
61 (m117)	GÖĞÜS HASTALIKLARI	Solunum Göğüs
62 (m119)	GÖĞÜS HASTALIKLARI	Solunum sistemi
63 (m121)	GÖĞÜS CERR.	Solunum sistemi
64 (m122)	GÖĞÜS CERR.	Solunum sistemi
65 (m123)	GASTROENTERO.	Sindirim sistemi
66 (m124)	GASTROENTERO.	Sindirim sistemi
67 (m125)	GASTROENTERO.	Sindirim sistemi
68 (m128)	ENDOKRİNOLOJİ	Endokrin
69 (m129)	ENDOKRİNOLOJİ	Endokrin
70 (m130)	ÜROLOJİ	Ürogenital sistem
71 (m131)	ÜROLOJİ	Ürogenital sistem
72 (m133)	NEFROLOJİ	Ürogenital sistem
73 (m139)	NÖROLOJİ	Sinir sistemi
74 (m140)	NÖROLOJİ	Sinir sistemi
75 (m141)	NÖROLOJİ	Sinir sistemi
76 (m142)	NÖROLOJİ	Sinir sistemi
77 (m145)	NÖROŞİRURJİ	Sinir sistemi
78 (m148)	PSİKİYATRİ	Mental, ruh sağlığı, psikiyatri
79 (m151)	PSİKİYATRİ	Mental, ruh sağlığı, psikiyatri
80 (m155)	ENFEKSİYON	Sindirim sistemi

Çizelge 8. 103 Maddelik Soru Bankasında Yer Alan Maddelerin Sistem ve Bilim Alanlarına Göre Dağılımı (Devamı)

Madde No	Bilim alanı	Sistem
81 (m156)	KADIN DOĞUM	Ürogenital sistem
82 (m157)	KADIN DOĞUM	Ürogenital sistem
83 (m158)	KADIN DOĞUM	Ürogenital sistem
84 (m160)	KADIN DOĞUM	Ürogenital sistem
85 (m161)	KADIN DOĞUM	Ürogenital sistem
86 (m165)	ÇOCUK SAĞLIĞI ve HAST.	Dolaşım
87 (m166)	ÇOCUK SAĞLIĞI ve HAST.	Duyu (göz, KBB)
88 (m171)	ÇOCUK CERRAHİ	Ürogenital sistem
89 (m172)	ÇOCUK CERRAHİ	Sindirim
90 (m177)	ORTOPEDİ	Kas iskelet
91 (m178)	ORTOPEDİ	Kas iskelet
92 (m180)	SPOR HEKİMLİĞİ	Kas iskelet
93 (m181)	SPOR HEKİMLİĞİ	Kas iskelet
94 (m182)	RADYOLOJİ	Ürogenital sistem
95 (m183)	ANESTEZİ	Dolaşım
96 (m185)	ANESTEZİ	Endokrin, metabolizma, beslenme
97 (m188)	ACİL	Kas iskelet sistemi
98 (m191)	GÖZ	Duyu (göz, KBB)
99 (m192)	GÖZ	Duyu (göz, KBB)
100 (m193)	KBB	Duyu (göz, KBB)
1001 (m194)	KBB	Duyu (göz, KBB)
102 (m195)	KBB	Duyu (göz, KBB)
103 (m200)	DERMATOLOJİ	Deri bağ Dokusu