T. C.
İstanbul Üniversitesi
Sosyal Bilimler Enstitüsü
İşletme Anabilim Dalı
Örgütsel Davranış Bilim Dalı

Doktora Tezi

ÇOK İHTİMALLİ RASCH MODELİ İLE DERECELENDİRİLMİŞ YANIT MODELİNİN ÖRTÜK ÖZELLİKLERİ TAHMİNLEME PERFORMANSI AÇISINDAN SİMÜLASYON YÖNTEMİYLE KARŞILAŞTIRILMASI

Serkan Dolma

2502030001

Tez Danışmanı
Prof. Dr. Erdal Tekarslan

İstanbul, 2009

ÖZ

Çok İhtimalli Rasch Modeli ile Derecelendirilmiş Yanıt Modelinin Örtük Özellikleri Tahminleme Performansı Açısından Simülasyon Yöntemiyle Karşılaştırılması

Serkan Dolma

Bu çalışmada Modern Test Kuramı çerçevesinde geliştirilen iki modelin örtük özellikleri tahminlemede ne kadar başarılı oldukları karşılaştırılmıştır. Buna ilaveten örneklem büyüklüğü, test uzunluğu ve şık sayısının yine tahminleme performansı üzerindeki etkileri incelenmiştir. Tahminleme performansı; yanlılık ve istikrar olmak üzere iki açıdan ele alınmıştır. Çalışma, simülasyon yöntemi ile üretilen suni veriler kullanılarak yapılmıştır. Elde edilen veriler, 'çok yönlü duble çokdeğişkenli varyans analizi' kullanılarak incelenmiştir.

Yapılan analiz sonucunda, Çok İhtimalli Rasch Modeli ile Derecelendirilmiş Yanıt Modeli arasında örtük özellikleri tahminleme başarısı açısından, önemli farklar olmadığı bulunmuştur. Dolayısıyla araştırmacıların tahminleme kalitesini arttırmak adına temel ölçme prensiplerinden vazgeçerek İki Parametreli Modellere yönelmelerine gerek olmadığı yorumu yapılmıştır. Ayrıca beklenildiği gibi, test uzunluğunun, şık sayısının ve örneklem büyüklüğünün artmasıyla tahminleme performansının arttığı sonucu elde edilmiştir. Buna ilaveten test uzunluğu ve şık sayısı arasında etkileşim olduğu sonucuna varılmıştır.

ABSTRACT

A Simulation Study for the Comparison of the Polytomous Rasch Model and Graded Response Model according to Their Performance on Recovering the Latent Traits

Serkan Dolma

In this study, two modern test theory models are compared in terms of their latent ability recovery performance. The effects of test length, category number and calibration sample size on trait estimations are also investigated. The performance of latent ability estimates are operationalized as two related but distinct concepts which are about the accuracy and consistency of the estimates. The analyses are conducted on artificial data which are produced according to the pre-specified simulation conditions. Data are then analyzed via multifactor doubly multivariate analysis of variance procedures.

The results of the analysis indicate that there is no significantly large difference between Polytomous Rasch Model and Graded Response Model in terms of ability estimation performance. Therefore it is concluded that researchers do not need to abandon fundamental measurement principles and apply two parameter models in order to increase model-data fit. It is also found that (as expected) increasing the number of items, categories and sample size have positive effect on the accuracy and the consistency of the estimates, and that there is an interaction between number of items and categories.

ÖNSÖZ

Bilimin üç temel yapıtaşı tasarım, ölçme ve analizdir. Bu öğeler etkileşim içerisindedir ve bir bilimsel çalışmanın sonucunda elde edilen verilerin ne kadar anlamlı olacağı bu üç unsurun ne kadar ciddi olarak ele alındığına bağlıdır. Ancak sosyal bilimlerde, bu üç unsurdan biri olan ölçmeye ne yazık ki gereken önem verilmemektedir. Elde edilen sonuçların, ölçmenin temel ilkelerine uygun değilse hiçbir anlamı olmadığı gerçeği ne yazık ki birçok araştırmacının gözünden kaçmaktadır. Araştırmacılar ve bilim felsefecileri arasında, ölçme mevzusuna gereken önemi göstermeyen ve ölçmeyi "nesnelerin özelliklerine sayısal değerler atamadan" ibaret gören bir kesim olduğu gibi, sosyal bilimlerde sayısal ölçümlerin yapılabilmesinin mümkün olmadığını düşünen bir kesimde bulunmaktadır. Bu iki kutbun arasında bir konumda bulunan ve sosyal bilimlerde ölçmenin belli prensiplere uyulması halinde mümkün olduğunu düşünenler ise çoğunluktadır. Böyle düşünenlerin geliştirdiği ölçme teorilerinden birisi de Madde Yanıt Kuramıdır.

Madde Yanıt Kuramı, bireylerin sorulara (ifadelere) belirli bir şekilde yanıt verme olasılıklarının, bireylerin yetenekleri ile ifadelerin bazı özelliklerinin bir fonksiyonu olduğu prensibine dayanmaktadır. Hangi madde özelliklerinin yanıt verme olasılığı üzerinde etkili olduğu modelden modele değişmektedir. Madde parametresinin artması modeli daha esnek hale getirerek veri-model uyumunu arttırmasına karşın bunun, temel objektif ölçme prensiplerinden uzaklaşmak gibi ağır bir bedeli vardır.

Bu çalışmanın birinci amacı, veri-model uyumunu arttırmak adına bu bedeli ödemeye razı olan örgütsel davranış araştırmacılarını, buna gerek olmadığı hususunda ikna etmektir. Mevcut çalışma ile ayırt edicilik parametresi maddeler arasında değişkenlik gösterse bile, küçük örneklemler ve az sayıda ifade ile çalışıldığında örtük özellik tahminlemesi açısından Rasch Modelleri ile İki Parametreli Modeller arasında fark olmadığı gösterilmiştir. İki parametreli modeller daha esnek olmasına karşın, daha iyi sonuçlar veremeyebilmesinin nedeni, bu modellerin parametre tahminlemesi için büyük örneklemlere ihtiyaç duymalarıdır. Örgütsel davranış araştırmaları dâhilinde yapılan ölçümler, genelde 150-300 arası büyüklükteki örneklemlerde yapıldığı için ve boyut başına düşen ifade sayısı az olduğu için iki parametreli modelleri kullanmaya hemen hiç gerek kalmamaktadır.

Çalışmada, her zaman Çok İhtimalli Rasch Modelinin tercih edilmesinin daha uygun olduğunu gösterildikten sonra şık sayısının ve ifade sayısının tahminleme üzerindeki etkileri analiz edilmiştir. Şık (kategori) sayısı ile ifade sayısı arasındaki dengenin nasıl olabileceği de grafiklerle gösterilmiştir.

Ayrıca örneklem büyüklüğü ile test uzunluğu ve tahminleme kalitesi arasında nasıl bir ilişki olduğu da incelenmiştir.

Çalışma boyunca bana her türlü serbestliği tanıyan danışman hocam Prof. Dr. Erdal Tekarslan'a, Winsteps programıyla ilgili yardımlarından dolayı Dr. Mike Linancre'a, araştırma tasarımıyla ilgili katkılarından dolayı Prof. Dr. Svend Kreiner'e, Rasch analizi ile ilgili yardımlarından dolayı Prof. Dr. David Andrich'e, veri girişi, kaynak temini gibi zahmetli işlerde bana yardım etmek için oldukça çok vakit ayıran arkadaşım Ebru Demirci'ye teşekkürlerimi sunarım.

İÇİNDEKİLER

ÖZ iii
ABSTRACT iv
ÖNSÖZ v
İÇİNDEKİLER vii
TABLOLAR LİSTESİ ix
ŞEKİLLER LİSTESİ x
1. GİRİŞ 1
1.1 Araştırma Sorusunun Arka Planı: Model-Veri Uyumu Meselesi 1
1.2. Araştırma Sorusu 3
1.3 Araştırmanın Gerekliliği 4
2. KONUYA İLİŞKİN LİTERATÜRÜN GÖZDEN GEÇİRİLMESİ 5
2.1 Madde Yanıt Kuramına Genel Bakış 6
2.2. İki İhtimalli Lojistik Modeller 8
2.3 Çok İhtimalli Modeller 16
2.3.1 Çok İhtimalli Rasch Modeli 17
2.3.2 Derecelendirilmiş Yanıt Modeli 22
2.4 Parametre Tahminleme Yöntemleri 24
2.5 Madde Yanıt Kuramının Genel Varsayımları 28
3. ARAŞTIRMA YÖNTEM BİLİMİ VE İZLENEN PROSEDÜRLER 30
3.1 Araştırmanın Tasarımına Genel Bakış 30
3.2 Araştırmanın Bağımlı Değişkenleri ve Araştırmanın Analiz Birimi 31
3.3 Araştırmanın Bağımsız Değişkenleri 35
3.4 Araştırmanın Kontrol Değişkenleri ve Araştırmanın Sabitleri 38
3.5 Araştırmanın Maddî Hipotezleri 39
3.6 Araştırmanın Sınırlılıkları 42
3.7 Simülasyon Yöntemi ve Veri Üretimi 45
3.8 Araştırmada Kullanılan Veri Analiz Yöntemi 46
3.9 Araştırma Sürecinin Adımları (Araştırmanın Akış Planı) 50
4. VERİ ANALİZİ VE BULGULAR 52

4.1 Varsayımların Kontrol Edilmesi 52
4.1.1 Madde Yanıt Kuramı Modellerine İlişkin Varsayımın Kontrolü 52
4.1.2 Veri Analiz Yöntemine İlişkin Varsayımların Kontrolü 55
4.2 Analiz Sonuçları 57
5. SONUÇ, GENEL DEĞERLENDİRME VE İLERİKİ ÇALIŞMALAR İÇİN ÖNERİLER ... 77
Ekler 80
Ek 1: Tezde Kullanılan Bazı Kavramların İngilizce Orijinalleri 80
Ek 2: Farksızlık Hipotezi İstatistiksel Anlamlılık Testleri (FYİAT) ile İlgili Açıklama 82
EK 3: Tablolar 84
KAYNAKÇA 89

TABLOLAR LİSTESİ

Tablo 4.1. TBA Sonucu Elde Edilen Açıklanan Varyans Yüzde ve Değerleri.... 55

Tablo 4.2. Çok Değişkenli Testler (Tüm Düzeyler için)..... 59

Tablo 4.3. Çok Değişkenli Testler (İki İhtimalli Modeller Hariç)..... 63

Tablo 4.4. Çok Değişkenli Testler (Çok Değişkenli Karma Model)..... 70

Tablo 4.5. Tek Değişkenli Testler..... 72

Tablo 4.6. Test Uzunluğu ile İlgili Polinom Planlı Mukayeseleri..... 76

ŞEKİLLER LİSTESİ

Şekil 3.1. Yeniden Kodlama Çizelgesi..... 44

Şekil 4.1. Ardışık Özdeğerler Grafiği..... 54

Şekil 4.2. Tahminleme Yanlılığına İlişkin Model x Kategori Sayısı Profil Grafiği..... 60

Şekil 4.3. Tahminleme Yanlılığına İlişkin Kategori Sayısı x Ölçek Uzunluğu Profil Grafiği..... 61

Şekil 4.4. Görgül Hata Varyansına İlişkin Model x Kategori Sayısı Profil Grafiği..... 62

Şekil 4.5. Görgül Hata Varyansına İlişkin Kategori Sayısı x Ölçek Uzunluğu Profil Grafiği..... 64

Şekil 4.6. Tahminleme Yanlılığına İlişkin Model x Kalibrasyon Örneklemi Büyüklüğü Profil Grafiği..... 66

Şekil 4.7. Görgül Hata Varyansına İlişkin Model x Kalibrasyon Örneklemi Büyüklüğü Profil Grafiği..... 66

Şekil 4.8. Tahminleme Yanlılığına İlişkin Kategori Sayısı x Ölçek Uzunluğu Profil Grafiği..... 67

Şekil 4.9. Görgül Hata Varyansına İlişkin Kategori Sayısı x Ölçek Uzunluğu Profil Grafiği..... 68

1. GİRİŞ

1.1 Araştırma Sorusunun Arka Planı: Model-Veri Uyumu Meselesi

Klasik Test Kuramının eksiklerine çözüm olması amacıyla 1950'lerin sonuna doğru şu an Modern Test Kuramı olarak anılan ve birçok varyasyonu olan bir yaklaşım geliştirilmeye başlanmıştır. Bu kuram çerçevesinde ele alınan birçok alt model bulunmaktadır ve bunların en belirgin ortak özelliği ölçme olgusunu olasılıksal bir çerçevede inceliyor olmalarıdır. Bu modeller, bir sorunun bazı özellikleri (zorluk seviyesi gibi) ile kişinin örtük özellik düzeyinin (mesela ilgili konudaki yeteneğinin), kişinin o soruyu belirli bir şekilde (mesela doğru olarak) yanıtlama ihtimalini belirleyeceği ilkesi üzerine kurulmuştur. Bu ortak özelliğe ve modellerin matematiksel olarak birbirlerine olan benzerliğine rağmen felsefi açıdan bazı modeller arasında çok ciddi farklar bulunmaktadır (Andrich, 2002). Ölçmenin tanımına, anlamına ve bilimsel çalışmalar açısından işlevine dair bu bakış açısı farklılığı, görünüşteki benzerliğe rağmen iki farklı görüş kampının oluşmasına neden olmuştur. Birinci yaklaşım olan Madde Yanıt Kuramı, model ile veriler arasındaki ilişkide ölçüm sonucu elde edilen verilerin temel olarak alınması gerektiğini, yani mevcut verilere göre eldeki en açıklayıcı modelin seçilmesi gerektiğini savunmaktadır. Bu yaklaşıma göre daha basit olan modeli kullanabilmek bir avantaj olarak görülmesine karşın veriler ile model arasında bir uyumsuzluk olması halinde bu avantajdan feragat edilip modelde, ilave parametreler ekleyerek, gerekli modifikasyonlar yapılmalı ve veri-model uyumu sağlanmalıdır.

Alternatif yaklaşımı temsil eden Rasch Modeli[1] yaklaşımı ise objektif, kişilerden ve sorulardan bağımsız bir ölçüm yapabilmek için verilerin bazı özelliklere sahip olması, yani verilerin modele uyması gerektiğini kabul etmektedir. Rasch bakış açısına göre ölçme girişiminde en önemli unsur koşullara göre değişiklik göstermeyen mukayeseler yapabilme ilkesidir. Bir başka değişle kişilerin (birbirlerine göre) nispi

[1] Her ne kadar Rasch Modeli savunucuları karşı çıksa da, 'Madde Yanıt Kuramı' başlığı aslında Rasch Modellerini de kapsayan genel bir isim olarak kullanılmaktadır. Bu giriş bölümü çerçevesinde, her seferinde uzun uzun yazmamak adına Madde Yanıt Kuramı kavramı ile "Rasch Modeli dışındaki diğer Madde Yanıt Kuramları" kastedilmektedir.

yeteneklerine ilişkin yapılan tahminler kullanılan sorulara göre değişkenlik göstermemeli, aynı şekilde soruların nispi zorluk düzeylerine ilişkin yapılan tahminler de ölçüm yapılan kişilere göre değişkenlik göstermemelidir. Dahası Rasch Modeli felsefesine göre bu özellik modelin değil verilerin sağlaması gereken bir özelliktir. Buna göre ölçüm sonucunda elde edilen veriler, model ile uyumsuzluk gösteriyorsa bu modelde değil, verilerin kendisinde bir sorun olduğu şeklinde yorumlanır. Dolayısıyla modelin mantıksal alt yapısından şüphe edilmiyorsa ve veriler model ile uyum göstermiyorsa bu, verilerin asgari bazı özellikleri sağlayamadığı, dolayısıyla bu haliyle işe yarar bir yanının olmadığı şeklinde kabul edilir. Örneğin bu yaklaşıma göre, aynı özelliğe yönelik bir ölçüm yapılmasına rağmen kullanılan sorulara göre kişilerin, sahip oldukları yetenek (özellik) düzeyleri açısından birbirlerine göre olan konumları değişiyorsa bunun en olasılı izahı aynı şeyi ölçtüğü düşünülen o soruların aslında başka şeyleri de ölçüyor olabileceğidir. Ayrıca Rasch Modelinde aynı özelliği ölçen soruların nispi zorluk düzeylerinin sorulara tabi tutulan kişilere göre değişkenlik göstermesi sorularda bir sorun olduğu şeklinde yorumlanır.

Madde Yanıt Kuramı (MYK) ise böyle bir durumda soruların ve verilen yanıtların olduğu gibi kabul edilip, modelin gerektiği şekilde "eğilip bükülmesini" tavsiye etmektedir. Bu yaklaşımın, alternatifine göre maliyeti oldukça az olmasına karşın, eğer objektif ve koşullara göre değişmeyen bir ölçüm yapılmak isteniyorsa bu yönde çok faydalı olmayacağı da bir gerçektir. İstenilen sonuçlar elde edilmeyince masa başında modeli değiştirmek ve verilere uydurmak elbette çok pratiktir, ancak bu ölçüm değerleri ile yapılacak analizlere ne derece de itimat edilmesi gerektiği tartışmalı hale gelecektir.

Rasch yaklaşımının altında yatan felsefenin ve ölçme olgusuna yönelik olan bakış açısının, Madde Yanıt Kuramınınkine oranla neden daha sağlam olduğu ile ilgili onlarca makale, bildiri ve yayın telif edilmiştir. Araştırmacıların pragmatik kaygılarla temel ölçme prensiplerinden ayrılıp diğer modellere yönelmemelerini telkin eden hayli esere rağmen, Madde Yanıt Kuramının elindeki çok güçlü koz (pratiklik) bu yaklaşımın yaygın kabul edilen bakış açısı ve hareket tarzı olarak yerleşmesine neden olmuştur. Bu durum, Rasch modelinin ise (çoğunluğa nazaran) radikal bir yaklaşım olarak kabul edilmesine sebep olmuştur. Madde Yanıt Kuramını temel almanın ne gibi sakıncaları olduğuyla ve model seçiminde aslında nelerin kriter olması gerektiği ile

ilgili bu güne kadar birçok yayın, makale hazırlanmıştır (mesela, Andrich, 2002 ve 2004; Wright, 1992 ve 1999; ve Wilson, 2002, bunlar arasında sıralanabilir). Dolayısıyla Rasch Modeli savunucularının yazdığı ve Madde Yanıt Kuramı bakış açısını benimsemenin mantıksal sonuçlarının incelendiği literatüre ilave edilebilecek pek bir şey kalmadığı söylenebilir. Bir başka değişle ölçmenin nasıl olması gerektiği konusunda ve ölçme faaliyetine nasıl yaklaşılması gerektiği hususunda yeterince ikna edici yayın bulunmaktadır. Ancak araştırmacıları (özellikle örgütsel davranış araştırmacılarını) belirli standartları olan kaliteli ölçümlere yöneltebilmek için, mantıksal argümanlar ileri sürmek yerine konuya başka bir açıdan yaklaşılabilir:

Madde Yanıt Kuramının en temel dayanak noktası, veri-merkezli bir felsefesi olduğu için bu kuramın verileri daha iyi açıkladığı gerçeğidir. Ancak gerçekte ölçme girişimcilerinin asıl ilgilendiği nokta merak edilen beşeri özelliğin mümkün olduğunca isabetli bir şekilde tahminlenmesidir. Gerçek insanlar ile yapılan ölçümlerde, gizli özelliğin doğru (gerçek) değerleri bilinmediğinden, farklı modellerin (ve farklı test koşullarının) tahminleme başarısı üzerindeki etkisi, ancak simülasyon çalışmalarıyla sınanabilmektedir. Madde Yanıt Kuramının veri-merkezli yapısı simülasyon çalışmalarına oldukça uyumlu olduğundan, bu tür çalışmalarda MYK modellerinin Rasch modellerine göre daha iyi sonuçlar vermesini beklemek gayet doğaldır. Bununla beraber MYK modelleri, isabetli ve istikrarlı tahminlemeler üretebilmesi için Rasch Modellerine göre daha fazla veriye (yani daha çok sorudan oluşan bir teste ve daha çok kişiden oluşan bir test grubuna) ihtiyaç duymaktadır. Mesela, yeterli örneklem büyüklüklerine erişilememesi durumunda bu model kullanılarak elde edilen sonuçların istikrarsız ve hatalı olma ihtimali bulunmaktadır. Örgütsel davranış araştırmalarında ise genelde, özellikle eğitim bilimlerine göre oldukça kısıtlı örneklemlerle çalışılmakta, merak edilen özellikler de katılımcıların vakitlerini çalmamak adına mümkün olduğunca kısa ölçüm araçları kullanılarak yapılmaktadır. Dolayısıyla bu gibi koşullar altında yapılan çalışmalar için Madde Yanıt Kuramının "sözde" avantajları da ortadan kalkmaktadır.

1.2. Araştırma Sorusu

Bu araştırmada cevaplanmaya çalışılan temel soru "kısa ölçekler ve küçük örneklemelerle yapılan ölçümlerde kişilerin örtük özelliklerinin tahminlenmesinde

kullanılan çok ihtimalli yanıt modeli türünün, tahminleme başarısı üzerinde önemli bir etkisi olup olmadığı" sorudur. Yani, bu araştırmanın temel güdüsü, tercih edilen ölçme modelinin, örtük özelliklere yönelik tahminlemeleri ne kadar etkilediği konusuna ilişkin meraktır. Daha spesifik hale getirilecek olursa, bu araştırmada birincil olarak, Rasch Modeli yerine diğer Madde Yanıt Kuramlarını tercih etmenin araştırmacıya doğru tahminlerde bulunmak açısından sağladığı bir avantaj olup olmadığı merak edilmektedir.

Bu sorunun yanı sıra parametre tahminleme performansı açısından modeller arasındaki farkın (ya da benzerliğin):

a) farklı adetlerdeki ifadelerden oluşan ölçekler için;

b) farklı şık sayılarından oluşan ölçekler için ve

c) farklı büyüklüklerde yapılan ölçüm seansları için de geçerli olup olmadığı incelenmek istenmektedir.

Ayrıca gerek, Rasch Modeli, gerekse Madde Yanıt Kuramı çerçevesinde sağlıklı ölçüm sonuçları alınabilmesi için gereken minimum ifade sayısı - kategori sayısı kombinasyonun nasıl olması gerektiği sorusuna cevap aranmaya çalışılacaktır. Bunun için bahsedilen faktörler arasında nasıl bir etkileşim olduğu incelenecektir.

1.3 Araştırmanın Gerekliliği

Madde Yanıt Kuramı çerçevesinde geliştirilen Çok İhtimalli Modelleri birbiriyle karşılaştırmak veya belli bazı modellere göre yapılan parametre tahminlemelerinin test koşullarından ne şekilde etkilendiğini tespit etmek amacıyla birçok araştırma yapılmıştır. Bu çalışmalar, tahminleme performansını nasıl ölçüldükleri, bağımsız değişken olarak neleri seçtikleri, kişi parametrelerinin mi yoksa madde parametrelerinin mi tahminlenme kalitelerini inceledikleri ve benzeri özellikleri bakımından farklılıklar gösterebilmektedir. Bu çalışmalara genel olarak bakıldığında şunlar söylenebilir:

(a) Çoğu zaman ölçme girişimi ile asıl amaçlanan kişilerin örtük özelliklerinin hangi koşullarda ve hangi modellerle daha iyi tahminlenebileceğini tespit etmek olmasına rağmen birçok çalışmada ya sadece ya da ağırlıklı olarak madde parametrelerinin tahminleme performansları incelenmiştir (mesela, de Ayala ve Sava-Bolesta, 1999). Test havuzu oluşturulması gibi özel durumların dışında çoğu zaman

asıl ilgilenilenin kişi özellikleri olmasına rağmen bağımlı değişkenleri bu doğrultuda oluşturulan araştırmaların sayısı oldukça azdır.

(b) Choi, Cook ve Dodd'un çalışması gibi istisnalar dışında (1997) birçok araştırmada kategori sayısı ve ölçek uzunluğu ayrı değişkenler olarak değil bir arada (toplam tahminlenecek parametre sayısı olarak) ele alınmıştır. Bir başka değişle bu iki faktör arasında etkileşim olmadığı farz edilmiştir.

(c) Birçok araştırmada çıkarsamacı (sonuç çıkarıcı) istatistiksel analizler uygulanılmamış, tasvir edici istatistikler ile yetinilmiştir (mesela, French ve Dodd, 1999).

(d) Yapılan araştırmaların çoğunda büyük kalibrasyon örneklemleri üzerinde durulmuştur. Ancak örgütsel davranış araştırmalarında ekseriyetle orta büyüklükteki örneklemler ile çalışıldığı için örgütsel davranışçılar bu araştırmaların sonuçlarından yeterince faydalanamamaktadırlar.

Literatür taraması sonucunda, çalışmaların yukarıda bahsedilen özelliklerinden dolayı, Modern Test Kuramı çerçevesinde ölçekler geliştirirken örgütsel davranış araştırmacılarına katkı sağlayabilecek, yol gösterici nitelikte yayın sayısının oldukça az olduğu belirlenmiştir. Buradan çıkışla, tipik örgütsel davranış araştırma koşullarını inceleyen bir simülasyon çalışmasının bu konudaki ihtiyacı kapatmada bir katkısı olabileceği kanısına varılmıştır.

Modern Test Kuramı dâhilinde ölçek hazırlayan araştırmacılara, örneklem büyüklüğü, ölçek uzunluğu ve şık sayısı gibi hususlarda karar verirken faydalanabilecekleri bazı bilgiler sunmak önemli olmasına karşın bu, çalışmanın ikincil hedefini oluşturmaktadır. Bu çalışmayı yapmaktaki asıl gaye, Madde Yanıt Kuramının Rasch Modellerine karşı sahip olduğu öne sürülen görünürdeki üstünlüklerinin, örgütsel davranış açısından istisnai durumlar dışında geçerli olmadığı, çoğu örgütsel davranış araştırması için Madde Yanıt Kuramını tercih etmenin görünürde bile bir avantajı olmadığını göstermektir.

2. KONUYA İLİŞKİN LİTERATÜRÜN GÖZDEN GEÇİRİLMESİ

Bu bölümde Rasch Modeli ailesi ve diğer Madde Yanıt Kuramı Modelleri kısaca incelenecektir. Madde Yanıt Kuramı başlığı altında yer alan bazı modeller, iki ihtimalli modellerden başlanarak gözden geçirilecektir. Çok ihtimalli modellerin de

incelenmesinden sonra, parametre tahminleme yöntemleri ve Madde Yanıt Kuramının genel varsayımları üzerinde kısaca durulacaktır.

2.1 Madde Yanıt Kuramına Genel Bakış

Madde Yanıt Kuramı (MYK), bireyin bir soruya belli bir şekilde yanıt verme olasılığının, kişinin sahip olduğu ve doğrudan gözlemlenemeyen özellik düzeyi ile sorunun zorluk düzeyinin (ve duruma göre, ayırt edicilik gücü ve sözde şans faktörü gibi sorunun diğer özelliklerinin) bir fonksiyonu olduğunu varsaymaktadır. Burada kişinin sahip olduğu özellik ile ne kastedildiği, ölçüm aracının türüne, tipine ve amacına göre değişebilmektedir. Mesela, belli bir yeteneği ölçmesi için hazırlanan bir testin soruları için 'kişinin özellik düzeyi' kavramı ile, bireyin o konudaki yetenek düzeyi (örneğin olaylar arası ilişki kurma, üç boyutlu düşünebilme yeteneği gibi) kastedilmektedir. Kişi özelliği kavramı, yetenek dışında, yerine göre; tutum, kişilik gibi başka soyut kavramlar için de kullanılabilir. Ancak, kuramın ilk çıkışı ve en yaygın kullanımı eğitim bilimlerinde olduğu için, bu kavram ile genelde kişinin belli bir konudaki yetenek düzeyi kastedilmektedir.

Madde Yanıt Kuramı[2] çerçevesinde birçok model geliştirilmiştir. Bu modeller, farklı açılardan çeşitli sınıflandırmalara tâbi tutulabilir. Mesela modele dâhil olan parametre sayısı bakımından ele alındığında Tek, İki, Üç ve Dört Parametreli Modeller olduğu söylenebilir. Bu modellerin birbirlerinden farkları bir sonraki bölümde kısaca incelenecektir.

Bağımlı değişken ile bağımsız değişken arasındaki ilişkiyi modellerken başvurulan matematiksel fonksiyon açısından 'Normal Kemer Modeller' ve 'Lojistik Modeller' olmak üzere ikili bir ayrım yapılmaktadır. Tarihsel açıdan önemi olmasına ve MYK'nın Klasik Test Kuramı (KTK) ile ilişkilendirilebilmesini sağlamasına karşın Normal Kemer Modeller, günümüzde pek tercih edilmemektedir[3]. Kavranılması, hesaplanması ve "cebirsel yönünün takip edilmesi" çok daha kolay olduğu için

---

[2] Madde Yanıt Kuramı yerine farklı isimler de kullanılabilmektedir. Bu alternatif isimlendirmelerden MYK'dan sonra en yaygın olanı Örtülü Özellik Kuramıdır (ÖÖK).

[3] Bu çalışmaya, kullanım sıklığı giderek azalan normal kemer modeller dâhil edilmemiştir.

günümüzde genellikle lojistik modeller kullanılmaktadır. Normal kemer modellerden lojistik modellere geçilmeye başlandığı ilk yıllarda iki model arasındaki farkı (modele göre çizilen eğriler[4] açısından) en aza indirmek için *D* sabiti[5] lojistik fonksiyona dâhil edilmekteydi. Böylece Normal Kemer Model kullanılarak analiz edilen testler ile lojistik fonksiyon kullanılarak analiz edilen testlerin sonuçları daha sağlıklı karşılaştırılabilmekteydi. Ancak günümüzde istisnalar dışında böyle bir karşılaştırmaya gerek duyulmadığından *D* sabiti de ilgili fonksiyonlarda kullanılmamaya başlanmıştır.

Yanıt ihtimallerinin sayısı bakımından ise 'İki İhtimalli' ve 'Çok İhtimalli' Modeller şeklinde ikili bir ayrım söz konusudur ve bu modeller bir sonraki bölümde incelenecektir.

Boyut sayısı bakımından 'Tekboyutlu' ve 'Çokboyutlu Modeller' olarak bir sınıflandırma yapılmaktadır. Soruya doğru yanıt verme ihtimalini etkilediği farz edilen yetenek sayısı modelin tekboyutlu mu yoksa çokboyutlu mu olacağını belirleyecektir. Literatürde ağırlıklı olarak tekboyutlu modeller incelenmektedir. Çokboyutlu modellerin henüz gelişme aşamasında olması, çokboyutlu analiz yapabilen yazılımların henüz geliştirilmemiş olması ve en önemlisi çokboyutlu yapılara ihtiyaç duyulan durumların gerçekte oldukça nadir olması gibi nedenlerden dolayı bu çalışma tek boyutlu modeller ile sınırlandırılmıştır.

Bu çalışmanın temel amacı çok ihtimalli ifadelerin kullanıldığı durumlarda (mesela Likert tipi ifadelerden oluşan tutum ölçeklerinde olduğu gibi), Tek Parametreli (1PL) ve İki Parametreli (2PL) modellerin kişi özelliklerini tahminlemedeki performanslarını karşılaştırmak olduğu için özellikle çok ihtimalli modeller üzerinde durulacaktır. Çok ihtimalli modellerin temelini oluşturduğu için önce iki ihtimalli modeller kısaca incelenecektir.

---

[4] "Yetenek" bağımsız değişkeni ile "doğru yanıtlama olasılığı" bağımlı değişkeni arasındaki ilişkiyi tasvir eden bu eğri Madde Karakteristik Eğrisi (MKE) olarak adlandırılmaktadır. Bu eğri MYK'nın temelini oluşturmaktadır. Öyle ki bazı yazarlar tarafından MYK ismi yerine Madde Karakteristiği Eğrisi Modeli ismi kullanılmaktadır.

[5] *D* sabiti genelde 1,7'ye eşitlenir. Ancak daha hassas bir uyum için 1,701 olarak kabul edilmesi önerilmektedir (McDonald, 1999, s. 269).

2.2. İki İhtimalli Lojistik Modeller

Madde Yanıt Kuramı dâhilinde yer alan modeller, cevaplandırma seçenekleri açısından İki İhtimalli (İİM[6]) ve Çok İhtimalli (ÇİM) olarak ikiye ayrılır.[7] Bu çalışma Örgütsel Davranış bilimi dâhilinde yapılan ölçümlerle ilgili olduğundan ve davranışsal ölçümlerde (özellikle tutumların ölçümünde) büyük ölçüde 'Likert Tipi' ifadelerden oluşan ölçekler kullanıldığından daha ziyade çok ihtimalli modeller üzerinde durulacaktır. Ancak çok ihtimalli modeller, iki ihtimalli modellerin genişletilmiş versiyonları oldukları için önce İİM'in özelliklerinin incelenmesi gerekmektedir.

İki İhtimalli Modeller, en çok iki farklı yanıt ihtimalinin olduğu 'yetenek testleri' veya diğer psikometrik ölçümler için önem taşımaktadır. Ancak yetenek dışında tutum, kişilik, davranış eğilimi gibi başka olguların ölçümünde de bu modellere başvurulabilmektedir[8].

Daha önce belirtildiği gibi genel olarak Madde Yanıt Modelleri kişinin yeteneği veya başka bir özelliği (mesela belli bir konuda sahip olduğu tutumun derecesi) ile bir ifadeye belli bir şekilde yanıt verme (mesela doğru olarak yanıtlama)

---

[6] 'İki ihtimalli' ve 'çok ihtimalli' ifadeleri, İngilizcedeki sırası ile '*dichotomous*' ve '*polytomous*' kavramlarına karşılık olarak kullanılmıştır. 'Şık' veya 'seçenek' yerine ihtimal kavramının tercih edilmesinin nedeni şudur: Bir soru ikiden fazla şıktan oluşabilir. Mesela tipik bir bilgi sorusu beş şıktan oluşur. Ancak teste tâbi tutulan kişi için sadece iki farklı ihtimal söz konusudur. Soruyu ya doğru ya da yanlış olarak cevaplandıracaktır. Gelişmiş bazı modellerin kullanıldığı durumlar dışında kişinin hangi yanlış şıkkı işaretlediğinin bir anlamı olmayacaktır. Dolayısıyla böyle bir soru beş şıklı fakat iki ihtimalli olarak tanımlanır. Olası yanlış anlamalara engel olmak için bu çalışmada '*dichotomous*' ve '*polytomous*' kavramlarının Türkçe karşılıklarında şık veya yanıt yerine ihtimal kavramı kullanılmıştır.

[7] Bunların dışında bir de Sürekli Yanıt (Veri) Modellerinden bahsedilebilir. Ancak, örgütsel davranış ölçümlerinde, sürekli veri niteliğinde yanıtları olan ölçüm araçlarına hemen hemen hiç başvurulmadığından Sürekli Yanıt Modellerine bu çalışmada yer verilmeyecektir. Konu ile ilgili olarak Mellenberg'e (1994) bakılabilir.

[8] Tarihsel olarak bakıldığında Madde Yanıt Kuramının gelişiminde ağırlıklı olarak eğitim bilimcilerin katkısı olduğundan olsa gerek, ölçülen olgu farklı olsa bile jenerik isim olarak 'yetenek' kavramı sıkça kullanılmaktadır.

olasılığı arasındaki ilişkiyi[9] modeller. Bu ilişkiye, dolayısıyla bu ilişkiyi matematiksel olarak betimleyen eğrilerin (MKE) özelliklerine, ilişkin varsayımlar aynı zamanda modelin parametre sayısını belirlemektedir.

Dört Parametreli Modelde[10] (4 PL), standart olarak tüm modellerde olması gereken madde zorluk (veya konum) parametresinin yanı sıra üç parametre daha yer almaktadır. Bu parametrelerden bir tanesi maddenin ayırt edicilik gücü ile ilgili, diğeri sorunun tahmin ile doğru yanıtlanabilmesiyle ilgili ve sonuncusu da kişinin üst düzey yeteneğe sahip olmasına karşın soruyu (dikkatsizlik v.b nedenlerle) doğru yanıtlayamaması ile ilişkilidir. Bir başka deyişle model, maddenin (ya da madde karakteristiği eğrisinin) bu dört özelliği ve kişinin yetenek düzeyi biliniyorsa, kişinin o maddeyi doğru olarak yanıtlama olasılığının hesaplanabileceğini öngörmektedir. Kişinin yetenek/özellik düzeyi, literatürde genelde 'kişi (veya yetenek) parametresi' olarak isimlendirilmektedir. Ancak yetenek düzeyini modelin bir parametresi olarak değil, değişkeni olarak ele almak teknik olarak daha doğru bir yaklaşım olacaktır.[11]
Bu çerçevede 4 parametreli lojistik model[12] şu şekilde formülize edilir (Barton ve Lord, 1981):

$$P(x_j = 1|\theta, \alpha_j, \delta_j, \chi_j, \gamma_j) = \chi_j + (\gamma_j - \chi_j)\frac{\exp\left[D\alpha_j(\theta - \delta_j)\right]}{1 + \exp\left[D\alpha_j(\theta - \delta_j)\right]} \qquad (2.1)$$

ki burada;

---

[9] MYK modellerinde bağımlı değişken "olasılık" olduğundan ve mantığa aykırı 'Y' değerleri elde edilmesine izin vermemek için (mesela P > 1 veya P < 0 gibi) lojistik regresyonda olduğu gibi değişkenler arasındaki ilişkinin "eğrisel" olduğu varsayılır.

[10] Tipik uygulamanın aksine (Reise ve Waller, 2002'den esinlenerek) bu çalışmada parametre sayısı açısından modelleri incelemeye 4 Parametreli Lojistik (4 PL) Modelden başlanılmıştır.

[11] Literatürde, bu yaklaşımı destekleyen herhangi bir kaynak bulunamadığından fakirin bu görüşü her türlü eleştiriye açıktır.

[12] Literatürde 'lojistik model' adı kullanılmasına rağmen 3 parametreli ve 4 parametreli modeller, ilgili parametrelerin modele dâhil olmasıyla teknik olarak lojistik model olma özeliklerini kaybederler.

θ kişinin yetenek (veya hangi özelliği ölçülüyorsa onun) düzeyini,

δ madde zorluk parametresini,

α maddenin ayırt edicilik parametresini,

χ "sözde şans" (veya alt asimptot) parametresini,

γ "sözde dikkatsizlik" (veya üst asimptot) parametresini ve

D ise ölçek/birim denkleştirme faktörünü göstermektedir[13].

Eşitliğin sol tarafı ise kişinin o soruyu doğru yanıtlama ihtimalini göstermektedir ve çoğu zaman kısaca $P_i(\theta)$ veya P_j olarak da gösterilebilmektedir.

Yetenek veya özellik kavramı ile kişinin dolaylı olarak ölçülmeye çalışılan ve gözlemlenen ilgili davranışlarının altında yattığı kabul edilen fakat kendisi doğrudan gözlemlenemeyen soyut olgunun kastedildiği daha önce belirtilmişti. Yetenek veya kişi parametresi ile bu olgunun (sosyodavranışsal bir özellik) kişide ne derecede bulunduğu sayısal olarak ifade edilmeye çalışılır. Bir başka deyişle kişi parametresi kişinin sahip olduğu özellik düzeyini gösterir. Bu arka planda var olduğu kabul edilen soyut özellik (çoğunlukla 'örtük' veya 'gizli özellik' olarak isimlendirilir); yetenek, kişilik özelliği, tutum, davranış eğilimi gibi farklı özellik türlerinden birisi olabilir. Örnekler verilecek olursa; olaylar arası ilişki kurma, dört işlem yapabilme v.b. ***yetenekler***; zekâ, hafıza kapasitesi gibi ***performansa dayalı özellikler***; içe dönüklük-dışa dönüklük gibi ***kişilik özellikleri***, Avrupa Birliğine karşı tutum, iş tatmini v.b gibi belli objelere karşı sahip olunan ***tutumlar*** veya belirli bir partiye oy verme niyeti gibi bir konudaki ***davranış eğilimleri*** gibi çok çeşitli kavramlardan birisi söz konusu örtük özelliği oluşturabilir. Yeter ki ölçülmek istenen özellik, gözlemlenen davranışların arkasında yatan temel soyut neden olarak tanımlanabilsin[14]. Bu arka planda yattığı farz edilen özelliğin doğrusal bir yapıya sahip olduğu ve kişinin yeteneği arttıkça (yani, bu farazi doğruda sağa doğru gidildikçe) diğer parametreler aynı kalmak kaydı

---

[13] Ayrıca tez boyunca *j* indisi ifadeleri, *i* indisi bireyleri ve *h* indisi kategori (şık) veya eşikleri temsil edecek şekilde kullanılmıştır.

[14] Ancak şu hatırlanmalı ki inceleme konusu olan özellik bu kadar çok çeşitlilik gösterebilmesine rağmen genel bir isim olarak 'yetenek' adı kullanılmaktadır. Bu çalışmada da asıl olarak örgütsel tutumların ölçülmesi üzerinde durulmasına rağmen, ölçülen bir özellikten bahsedilirken 'yetenek' tabiri kullanılmıştır.

ile soruyu doğru yanıtlama olasılığının arttığı kabul edilir. Bu bireysel bir özelliği temsil eden doğru (yatay eksen) 'logit' denen birimle ölçeklendirilmiştir (Henard, 2000) ve teorik olarak $-\infty$ ile $+\infty$ arasında değişebilirken uygulamada genellikle yalnızca -3 ile +3 arası incelenir ve grafiklerde gösterilir. Bu çerçevede 0 (sıfır) logit düzeyinde bir yeteneği olan kişi orta düzeyde yetenekli anlamına gelmektedir.

Madde zorluk parametresi ise sorunun doğru olarak yanıtlanabilmesinin ne kadar güç olduğunu göstermektedir. Bu parametre yetenek parametresi ile aynı birime sahiptir. Bu belki de Madde Yanıt Kuramının en kritik özelliğidir. Yetenek ve zorluk parametrelerinin aynı birime sahip olmaları onları doğrudan mukayese edilebilir hale getirmektedir. Böylece iki parametre arasındaki farka göre kişinin ilgili soruyu yanıtlama olasılığı kestirilebilir. Örneğin, θ (kişinin yetenek düzeyi) +1,0 iken δ (sorunun zorluk düzeyi) +0,5 ise o kişinin soruyu doğru yanıtlama olasılığının, yanlış yanıtlama olasılığından daha fazla olduğu, hiçbir işlem yapmadan rahatlıkla söylenebilecektir.

Eğer yetenek veya bilgi değil de örneğin bir tutum ölçülüyorsa zorluk parametresi ile o ifadeyi doğru olarak yanıtlama olasılığı değil, ifadedeki görüşe katılmanın (aynı fikirde olmanın) ne kadar zor olduğu kastedilir. Mesela bir markaya olan sadakatin ölçüldüğü bir ölçekte "İlk tercihimi X markasından yana kullanmaya çalışırım" ifadesine göre "Mağazada X markası yoksa o markayı bulana kadar diğer mağazaları dolaşırım" ifadesiyle aynı fikirde olmak kişide *daha yüksek* marka sadakati olmasını gerektirecektir. Bir başka deyişle belirli bir düzeyde marka sadakati olan bir müşterinin, ilk ifadeye "Tamamen katılıyorum" şeklinde yanıt verme olasılığı ikinci ifadeye göre daha yüksektir.

Ayırt edicilik parametresi, bir sorunun daha yüksek yetenekliler ile daha düşük yeteneklileri birbirlerinden ne kadar keskin bir şekilde ayırt edebildiğini sayısal olarak gösterir. Bu parametre aynı zamanda yetenek düzeyi ile soruyu doğru yanıtlama olasılığı arasındaki ilişkiyi tasvir eden madde karakteristik eğrisinin büküm noktasındaki eğimine oranlıdır[15]. MKE'nin daha dik olması, o noktada sorunun ayırt ediciliğinin daha yüksek olduğunu, bir başka deyişle sorunun, yeteneği yüksek olan ile

---

[15] 2 PL modellerde büküm noktasında eğim a_j'nin 0,25 katıdır (de Ayala, 2009, s. 120).

düşük olanı daha keskin bir şekilde ayırt edebildiğini göstermektedir. Ayırt edicilik parametresi de teorik olarak -∞ ile +∞ arasında bir değer alabilir. Ancak negatif değerler amacının tersine işleyen maddeleri gösterecektir. Bu maddeler araştırmacı için bir anlam ifade etmez. Araştırmacılar negatif değerlerin yanı sıra, pozitif ama 1'den çok düşük ayırt edicilik parametresine sahip ifadeleri de ölçeklerinde kullanmak istemeyeceklerdir. Çünkü bu tip ifadelerin kişilerin özelliklerini tahminlemedeki katkısı oldukça düşük olacaktır. Aşırı yüksek değere sahip (mesela 3'den büyük) ifadeler ise yetenek uzanımı[16] üzerinde belirli bir noktada mükemmele yakın bir hassaslıkta ayırım yapabilirken o noktadan uzaklaştıkça ayırt edicilik özelliklerini kaybederler. Örneğin, ayırt edicilik parametresi 5 olan ve zorluk parametresi +2,0 olan bir soru +1,7 ile +2,3 arasında yeteneğe sahip kişileri birbirlerinden çok hassas bir şekilde ayırt edebilirken -2 civarında yeteneğe sahip kişileri birbirlerinden ayırt etmede[17] pek katkı sağlamayacaktır.[18]

Genel olarak yetenek düzeyi azaldıkça bir soruyu doğru yanıtlama olasılığı azaldığından ve teorik olarak yetenek düzeyinin alt sınırı olmadığından madde karakteristik eğrisinin x-eksenine asimptotik olacağı düşünülür. Ancak bazı durumlarda yetenek ne kadar azalırsa azalsın soruyu doğru yanıtlama olasılığı sıfıra yaklaşmaz. Örneğin beş şıktan oluşan çoktan seçmeli bir test sorusuna cevap veren bir kişinin yeteneği ne kadar az olursa olsun doğru yanıt verme olasılığı en az 0,20'dir[19]. "Sözde Şans" veya "Sözde Tahmin" parametresi ise madde karakteristik eğrisinin 'alt (doğrusal) asimptotunun' konumunu temsil etmektedir. Yukarıdaki örnek için MKE'nin alt asimptotu, teorik olarak, $y=P_i(\theta)=0,2$ doğrusudur. Dolayısıyla modeldeki χ_j parametresi 0,2 ye eşit olacaktır.

---

[16] Uzanım ifadesi Cemal Yıldırım'ı takiben (1991) İngilizcedeki '*continuum*' kavramına karşılık olarak kullanılmıştır.

[17] Bu örnek iki parametreli bir model için verilmiştir.

[18] Baker'ın (2001) İnternetten ücretsiz olarak edinilebilen elektronik kitabıyla beraber gelen Java uygulaması, parametreler ile MKE'lerinin formları arasındaki ilişkiyi görsel uygulamalarla açıklamaktadır (erişim için, http://echo.edres.org:8080/irt/baker/).

[19] Tabiî bu durum, kişinin soruya mutlaka cevap vermek zorunda olduğu ve bilmediği durumlarda tahminde bulunduğu varsayılırsa geçerlidir.

Bazı durumlarda ise, beklenilenin aksine yetenek düzeyi ne kadar artarsa artsın, soruyu doğru yanıtlama olasılığı [$P_j(\theta)$] 1'e yaklaşmayabilir. Bunun farklı nedenleri olabilir. Mesela üstün yetenekli kişilerin sahip oldukları ilave bilgilerin soruyu hazırlayan kişinin standart olarak kabul ettiği doğrular ile çelişmesi veya üstün yetenekli kişilerin soruya gereken dikkati vermemeleri muhtemel nedenler arasında sayılabilir. Böyle bir durumda MKE'nin üst asimptotu y=$P_j(\theta)$=1 doğrusu olmayacaktır. 4 PL Modellerde, γ_j parametresi MKE'nin üst asimptotunun konumunu göstermektedir[20]. Ancak Barton ve Lord (1981) dördüncü parametrenin modele dâhil edilmesinin pratik hiçbir faydası olmadığını belirtmişlerdir. Gerçekten de Dört Parametreli model hemen hiç kullanım alanı bulamamıştır.

γ_j'nın 1'e eşitlenmesiyle model üç parametreli (Birnbaum, 1968) hale döner ve şu şekilde formülleştirilir[21]:

$$P(x_j = 1|\theta, \alpha_j, \delta_j, \chi_j) = \chi_j + (1 - \chi_j)\frac{\exp\left[\alpha_j(\theta - \delta_j)\right]}{1 + \exp\left[\alpha_j(\theta - \delta_j)\right]} \qquad (2.2)$$

Alt asimptot parametresinin modelde yer alması, model-veri uyumunu arttırmasına rağmen birçok dezavantajı da beraberinde getirir. Bunlardan, bazı durumlarda parametre tahminlenmesinde kararlı sonuçlara ulaşılamaması (Linacre, 2004a) ve gerekli örneklem büyüklüğünü misli ile arttırması en önemlileridir. Buna rağmen eğitim bilimlerinde, özellikle çoktan seçmeli testlerin kitlesel uygulamalarında (örneğin lisans ve lisansüstü giriş sınavları) bu model sıkça kullanılmaktadır[22]. Öte yandan yetenek veya bilgi dışındaki sosyodavranışsal ölçümlerde alt asimptot

---

[20] Bu parametreler için kullanılan "sözde şans" ve "dikkatsizlik" isimleri yanlış bilgilendirmeye neden olabilmektedir. Çünkü bu isimler söz konusu parametrelerin madde ile ilgili değil de kişi ile ilgili olduğu izlenimini uyandırmaktadır. Böyle bir karışıklığa izin vermemek için teknik açıdan daha doğru olan 'alt asimptot parametresi' ve 'üst asimptot parametresi' ifadelerinin kullanılması daha isabetli olacaktır.

[21] Bundan sonraki modeller için D sabiti daha önce belirtilen gerekçe ile model denklemlerine dâhil edilmeyecektir.

[22] Ancak çoğu uygulamada alt asimptot parametresi tahminlenmez (değeri araştırmacı tarafından atanır) ve/veya testteki tüm sorular için eşit kabul edilir.

parametresinin varlık gerekçesi büyük ölçüde ortadan kalkmaktadır. Örneğin 'Örgütsel Bağlılığın' ölçülmesinde alt asimptot parametresinin mevcudiyetini açıklamak pek mümkün değildir[23]. χ_j parametresinin sıfıra eşitlenmesiyle model iki parametreli lojistik modele dönüşmektedir[24]:

$$P(x_j = 1|\theta, \alpha_j, \delta_j) = \frac{\exp\left[\alpha_j(\theta - \delta_j)\right]}{1 + \exp\left[\alpha_j(\theta - \delta_j)\right]} \quad (2.3)$$

Son olarak ayırt edicilik parametresinin bir ölçekteki tüm ifadeler için aynı olduğu durumlarda modeldeki α parametresinin j indisi düşer ve model tek parametreli lojistik model haline dönüşür:

$$P(x_j = 1|\theta, \alpha, \delta_j) = \frac{\exp\left[\alpha(\theta - \delta_j)\right]}{1 + \exp\left[\alpha(\theta - \delta_j)\right]} \quad (2.4)$$

α parametresi tüm ölçek için sadece bir kere tahminlendiği için, bir başka deyişle ifadeden ifadeye değişiklik göstermediği için diğer parametrelerden farklı bir anlam kazanır.[25] Eğer α=1 olarak kabul edilirse, gösterimi şu şekilde olacaktır:

$$P(x_j = 1|\theta, \delta_j) = \frac{\exp\left(\theta - \delta_j\right)}{1 + \exp\left(\theta - \delta_j\right)} \quad (2.5)$$

---

[23] Böyle bir ölçümde bu parametrenin modele dâhil edilmesi ancak ölçülen özellik bakımından en alt düzeyde olan kişinin en düşük seçeneği işaretlememesinin bir gerekçesi olduğu durumlarda (mesela anketlerin gizliliğinden şüphe edildiği için katılımcıların samimi cevaplar vermekten çekindiği bir durumda) anlam kazanabilir. Ancak uygulamada tutum ve benzeri olguların ölçülmesinde 3 PL model kullanılmamaktadır.

[24] Bazen bu denklemin cebirsel dengi olan $P(x_j = 1|\theta, \alpha_j, \delta_j) = \frac{1}{1+\exp\left[-\alpha_j(\theta-\delta_j)\right]}$ veya $P(x_j = 1|\theta, \alpha_j, \delta_j) = \{1 + \exp\left[-\alpha_j(\theta - \delta_j)\right]\}^{-1}$ da kullanılabilmektedir.

[25] Bir açıdan, "bu durumda α, modelin parametresi değil katsayısı olur" da denebilir.

Ayırt edicilik parametresini tüm ölçek için bir kere tahminlemek ile 1'e eşit kabul etmek arasında uygulamada herhangi bir fark yoktur. Yalnızca tahminlenen yetenek parametresinin birimleri farklı olduğu için sonuçlar sayısal olarak farklı olacaktır. Ancak iki model ile tahminlenen yetenek parametrelerinin korelasyonu 1'e çok yakın olacaktır.

Burada verilen denklem cevaplayıcının soruyu doğru yanıtlama olasılığını göstermektedir ki çoğu zaman ilgilenilen cevap ihtimali budur. Ancak yanlış şıkkı seçme olasılığı ile ilgilenecek olunursa, $P(x_j = 1) + P(x_j = 0) = 1$ olduğundan (iki ihtimalli durumlarda),

$$P(x_j = 0|\theta, \delta_j) = 1 - \frac{\exp(\theta - \delta_j)}{1 + \exp(\theta - \delta_j)} = \frac{1}{1 + \exp(\theta - \delta_j)} \quad (2.6)$$

olacaktır. Doğru ve yanlış cevap verme ihtimallerini birlikte kapsayacak genel bir denklem ise;

$$P(x_j = x|\theta, \delta_j) = \frac{\exp x_j(\theta - \delta_j)}{1 + \exp(\theta - \delta_j)} \quad (2.7)$$

şeklinde düzenlenecektir ki burada x_j işaretlenen şıkkı (İİM'de yalnızca x_j=0 veya x_j=1 olabilir) veya aşılan eşik[26] sayısını temsil etmektedir[27]. Eşik sayısı her zaman ihtimal sayısının bir eksiği olduğundan iki ihtimalli bir ölçümde yalnız bir eşik olacaktır.

Rasch modeli ile tek parametreli lojistik model isimleri literatürde genelde birbirlerinin yerine kullanılmaktadır.[28] Ancak Rasch modelinin parametrelendirilmesinde genelde (2.5) denklemi tercih edilmektedir.

---

[26] Bu genel denklem iki ihtimalli modeller için çok gerekli olmayabilir ancak iki ihtimalli modellerden çok ihtimalli modellere kuramsal olarak geçişte ve tahminleme yöntemlerinde bu genel denkleme ihtiyaç duyulacağından burada yer verilmiştir. Eşik kavramı, çok ihtimalli modeller incelenirken daha ayrıntılı olarak ele alınacaktır.

[27] Yukarı denklemlerdeki paydaların aslında, payın alabileceği tüm değerlerin toplamı olduğuna dikkat ediniz. Bir başka değişle, $1 + \exp(\theta - \delta_j) = \exp 0(\theta - \delta_j) + \exp 1(\theta - \delta_j)$'dir.

Yukarıdaki açıklamalardan da anlaşılacağı üzere örgütsel tutum veya benzeri olguları ölçmek isteyen bir araştırmacı esas olarak iki model arasında bir seçim yapmak durumundadır: Araştırmacı Rasch modelini temel alabilir veya iki Parametreli MYK'nı tercih edebilir. Ancak örgütsel ölçümler genelde çok ihtimalli ifadeler ile yapıldığından, araştırmacı aslında Rasch modeli ve 2 PL Modelin çok ihtimalli uzantıları arasında seçim yapacaktır. Bu modellerden bazıları bir sonraki bölümde incelenecektir.

2.3 Çok İhtimalli Modeller

İki ihtimalli ifadelerden oluşan ölçüm gereçleri birçok sosyodavranışsal ve örgütsel olgunun ölçümünde yetersiz kaldığından ve uygulamada çok ihtimalli ifadelerden oluşan ölçeklerin (özellikle Likert tipi ifadelerden[29] oluşan ölçekler) çok yaygın olarak kullanıldığından dolayı bu tarz ölçümleri MYK ve Rasch Modelleri bakış açısıyla modelleyebilme ihtiyacını ortaya çıkarmıştır[30]. Bu çerçevede birçok çok ihtimalli MYK modeli geliştirilmiş ve halen geliştirilmektedir. Bu çalışmada çok ihtimalli modellerin birbirlerine göre karşılaştırılmasından ziyade Rasch Model ailesi ile 2 PL Model ailesi karşılaştırdığından sadece iki model üzerinde durulacaktır. Bunlar Çok İhtimalli Rasch Modeli (Andrich 2008; Luo, 2005) ve Derecelendirilmiş Yanıt Modelidir (Samejima, 1969 ve 1997). Bunların dışında simülasyon ile veri üretiminde temel alınacağı için Muraki'nin (1992 ve 1997) Genelleştirilmiş Kısmî Puanlama Modeli'ne de kısaca değinilecektir.

---

[28] Cebirsel olarak tek parametreli madde yanıt kuramı ile Rasch modeli eşit olmasına karşın aralarında her şeyin anlamını değiştiren bir felsefe farkı vardır. Bu tezin çıkış noktalarından birisi de Rasch Modelinin felsefesinin üstünlüğüne dair sahip olunan kanıdır.

[29] Likert ölçeği (Likert, 1932) ve Likert tipi ifadelerden oluşan ölçek kavramları arasındaki fark için bkz. Uebersax (2006).

[30] Masters'ın (1982) modelinin çıkış noktası biraz daha farklıdır. Bu model, özellikle doğru-yanlış tarzı sorular için geliştirilmiştir. Ancak model, dereceleme ölçekleri ve tutum ölçekleri için de kullanılabilmektedir.

2.3.1 Çok İhtimalli Rasch Modeli

İki ihtimalli Rasch modellinin çok ihtimalli ölçümler için genelleştirilmiş hâli olan bu model, Rasch Modelinin tüm kritik özelliklerine sahiptir. Bir başka deyişle, ham skorlar tahminleme için yeterli istatistiktir, yetenek tahminleri toplanabilirlik özelliğini korurular ve parametrelerin 'ayrıştırılabilirlik' özellikleri (Rasch, 1960, s. 122, Masters, 1982) kaybolmamıştır. Ancak literatürde Çok İhtimalli Rasch Modeli (ÇİRM) isimli bir modelden genelde bahsedilmemektedir. Yerine, Dereceleme Ölçeği Modeli (Andrich, 1978 ve 2007) ve Kısmî Puanlama Modeli (Masters 1982 ve 1988) olarak iki ayrı modele değinilmektedir. Çıkış noktaları farklı olsa da iki modelin aslında aynı olduğu, dolayısıyla böyle bir ayrıma gerek olmadığı (Luo, 2005) için bu çalışmada genel adıyla Çok İhtimalli Rasch Modeli başlığı benimsenmiştir. Ancak literatürle paralellik oluşturmak amacıyla önce Andrich'in ve Masters'ın modelleri ayrı ayrı incelenecektir.

Dereceleme Ölçeği Modeli (DÖM) spesifik olarak Likert'in geliştirdiği yanıt formatını kullanan ifadelerden oluşan ölçekler için geliştirilmiş bir Rasch modelidir. Bilindiği gibi bu format tarzı ardışık şıklardan oluşmaktadır. Deneklerden[31], yoğunluğu (veya gücü, önemi, kıymeti v.b. bir özelliği açısından) bir sıra takip eden bu şıklardan kendi tutumunu (veya düşüncesini, görüşünü, kişiliğini v.b. özelliğini) en iyi temsil eden şıkkı işaretlemesi istenmektedir. Genelde 5, 4, 3 veya 7 şıktan oluşan ifadeler kullanılmaktadır. Klasik Test Kuramı çerçevesinde hazırlanan ölçeklerde bu şıkların genelde eşit aralıklı veri tipinde ölçüm sağladığı kabul edilir.[32] Dereceleme Ölçeği Modelinde ise bu şıkların yetenek/özellik uzanımını ardışık ancak eşit olmayan parçalara böldüğü varsayılır. Örneğin dört şıklı (ihtimalli) bir tutum ölçeği ifadesi, tutumun yoğunluğunu temsil eden uzanımı dört ardışık parçaya bölecektir. Bu farazi

---

[31] Bazı akademisyenler İngilizcedeki '*respondent*' veya '*subject*' kavramlarına karşılık olarak 'denek' kavramını insanlar için kullanmayı önermemektedirler. Ancak yerine önerilen 'katılımcı' kavramı daha çok karışıklığa neden olduğundan ve bu ifadede bir aşağılama olduğu düşünülmediğinden bu çalışmada söz konusu kavramlara karşılık olarak 'denek' kavramı kullanılmıştır.

[32] Bu aslında Klasik Test Kuramının bir varsayımı değildir. Ancak araştırmacılar, sıralı veriye nazaran sağladığı avantajları dikkate alarak bu tip ifadeleri eşit aralıklı ölçek düzeyinde kabul etme eğilimindedirler. Bu konuyla ilgili olarak bkz. Byrne (2001, s. 71-72).

bölme işlemini şık sayısından bir eksik sayıda olan eşiklerin gerçekleştirdiği kabul edilir. Dolaysıyla bu örnekte üç eşik (veya kritik nokta) söz konusudur. Bir deneğin sahip olduğu tutumun yoğunluğu birinci farazî eşiği aşacak kadar değilse, işaretlemesi en muhtemel yanıt ilk şıktır (tipik bir tutum ölçeğinde bu "Kesinlikle Katılmıyorum" etiketi ile gösterilir). Sahip olunan tutum birinci eşiği geçecek kadar yoğun (güçlü) ama ikinci eşiği aşacak kadar güçlü değilse seçilmesi en muhtemel şık ikinci şıktır. Aynı mantık diğer tüm şıklar için de geçerlidir. Buradaki beklenti bir eşiği geçen kişilerin, o eşikten evvelki eşikleri zaten geçeceği şeklindedir. Ancak görgül sonuçların bu beklenti ile uyuşup uyuşmadığı kontrol edilmelidir (Andrich, 1982). Yani beklentilerin aksine bir sorunun "katılıyorum" şıkkını işaretlemek, "kesinlikle katılıyorum" şıkkını işaretlemekten daha yoğun bir tutum sahibi olmayı gerektirebilir. Bu durum, ifade yazarı ile cevaplayıcı arasında ifadeyi anlama bakımından bir fark olduğunun işareti olarak kabul edilecektir. Araştırmacı bu durumun nedenlerini kalitatif açıdan incelemeli (Andrich, de Long ve Sheridan, 1997) ve gerekirse soruyu değiştirmeli veya ölçekten çıkarmalıdır.[33]

Bu modelde maddeler ile ilgili zorluk parametresine ilave olarak eşiklerin zorluk parametreleri de modele dâhil edilmiştir.[34] Cevaplayıcının belli bir şıkkı işaretleme olasılığı, o sorunun zorluk düzeyi ile o şıkkın (tüm ölçek için aynı olan) eşik zorluk değeri ile toplamına göre belli olacaktır. Yani cevaplayıcının bir şıkkı seçme olasılığı, yeteneği ile sadece sorunun zorluk düzeyi arasındaki farkın fonksiyonu [$P(x)=f(\theta-\delta_j)$] değil, yeteneği ile sorunun zorluk düzeyinin eşik değeri ile toplamının farkının fonksiyonudur [$P(x)=f(\theta-(\delta_j+\tau_h))$]. Bir başka deyişle bir ölçekteki iki ifadenin DÖM'ne göre eşik değerleri aynı olması gerekirken, eşiklerin konumlarının iki ifade için de aynı olabilmesi ancak ifadelerin zorluk düzeylerinin aynı olması durumunda söz konusunudur. Öte yandan konumları ifadeden ifadeye farklılık göstermesine rağmen ardışık iki eşik arasındaki fark tüm ifadeler için aynı olacaktır.

---

[33] Masters (1988, s. 23) ise böyle bir düzensiz şık yapısının olduğu gibi kabul edilebileceğini söylemektedir (Wright ve Masters, 1982, s. 42, Masters, 1988) ancak Andrich, de Long ve Sheridan'ın, (1997) izahı Rasch Modeli felsefesiyle daha tutarlıdır.

[34] Ancak Andrich'in (1978) ilk makalelerinde bu parametrelerin her ifade için aynı olacağı kabul edilmiştir.

Formülleştirilecek olursa model;

$$P(x_j|\theta, \delta_j, \tau_1, \tau_2, \ldots \tau_m) = \frac{\exp\{\sum_{h=0}^{x_j}[\theta - (\delta_j + \tau_h)]\}}{\sum_{k=0}^{m}\exp\{\sum_{h=0}^{k}[\theta - (\delta_j + \tau_h)]\}} \quad (2.8)$$

şeklinde gösterilebilir ki burada τ_1 , τ_2 ... τ_h ile sırasıyla eşik düzeyleri, m ile toplam eşik sayısı, h indisi ile eşik numarası sembolleştirilmiştir.[35] Yukarıdaki denklem, ifadenin tüm eşik değerlerinin bilinmesi (ki bu eşik değerleri ölçekteki tüm ifadeler için aynıdır), sorunun zorluk düzeyinin ve kişinin yetenek düzeyinin bilinmesi halinde kişinin belli bir şıkkı işaretleme olasılığının ne olacağını belirtmektedir.[36] Eşik değerleri, sorunun zorluk düzeyinden sapmaları temsil ettiği için, modelde, eşik değerleri $\sum_{h=0}^{m}\tau_h = 0$ olacak şekilde kısıtlanmıştır.

Yukarıdaki denklemde h indisi olmayan elemanlar toplam sembolünün dışına çıkarılırsa ve eşitliğin sol tarafını kısaca $P(x_j)$ olarak gösterilirse denklem,

$$P(x_j) = \frac{\exp[(-\sum_{h=0}^{x_j}\tau_h) + x_j(\theta - \delta_j)]}{\sum_{k=0}^{m}\exp[(-\sum_{h=0}^{k}\tau_h) + \mathrm{k}(\theta - \delta_j)]} \quad (2.9)$$

haline dönüşecektir (de Ayala, 2009). Buradaki $-\sum_{h=0}^{x_j}\tau_h$ elemanı, cevaplayıcının en son geçtiği eşiğe kadar olan eşik değerlerinin toplamının -1 ile çarpımını temsil etmektedir. Bu, κ_{x_j} ile gösterilecek olursa denklem;

---

[35] Burada m_j yerine m kullanılması, şık (ve eşik) sayısının tüm ifadelerde aynı olması gerektiğini gösterirken τ_{jh} yerine τ_h kullanılması eşik değerlerinin ölçekteki tüm maddeler için aynı olduğunu ifade etmektedir. Ayrıca de Ayala (2009) yalnızca şık sayısının değil şıkların işleyiş şeklinin de tüm ifadeler için aynı olması gerektiğini belirtmiştir. Bu gerekliliğe kabaca 'eş etiketler koşulu' denebilir.

[36] Buradaki x_j kişinin kaçıncı şıkkı işaretlediğini değil kaçıncı eşiği geçtiğini temsil etmektedir. Dolayısıyla x_j 0 ile m arasında bir değer alabilir. x_j=0 olması kişinin en alt eşiği bile geçemediğini yani muhtemelen ilk şıkkı (genelde bu "kesinlikle katılmıyorum" şıkkıdır) işaretleyeceğini ifade etmektedir.

$$\left(x_j \middle| \theta, \delta_j, \kappa_{0_j}, \kappa_{1_j}, \ldots\ \kappa_{m_j}\right) = \frac{\exp\left[\kappa_{x_j} + x_j(\theta - \delta_j)\right]}{\sum_{k=0}^{m} \exp\left[\kappa_{k+} k(\theta - \delta_j)\right]} \qquad (2.10)$$

haline dönüşecektir. Buradaki κ_{x_j} terimi[37] 'kategori katsayısı' olarak (Andrich, 1978), isimlendirilmektedir. Dikkat edilirse çok ihtimalli ölçümler için geçerli olan bu denklem ile (2.7) denklemi arasındaki büyük bir benzerlik vardır. Aslında bu denklem (2.7)'in genelleştirilmiş hâlidir.[38] İhtimal sayısı iki olduğunda iki denklem aynı değerleri verecektir.

Kısmî Puanlama Modelinde ise eşik değerlerinin tüm ifadeler için aynı olması gerektiği kısıtı serbest bırakılmıştır. Ancak Masters (1982, 1988) modelini DÖM'ün genelleştirilmesi şeklinde oluşturmamıştır. Modelde, şıkkın soru içindeki zorluğunu temsil eden (τ_{jh}) ve sorunun test içindeki zorluğunu temsil eden (δ_j) olmak üzere iki ayrı parametre kullanılması yerine tek bir parametre (δ_{jh}) kullanılmıştır. Bu haliyle model;

$$P\left(x_j \middle| \theta, \delta_{hj}\right) = \frac{\exp\left[\sum_{h=0}^{x_j}(\theta - \delta_{jh})\right]}{\sum_{k=0}^{m_j} \exp\left[\sum_{h=0}^{k}(\theta - \delta_{jh})\right]} \qquad (2.11)$$

şeklinde gösterilebilir[39] ki burada x_j aşılan basamak sayısını göstermektedir.

Bu şekilde parametrelendirilmesiyle, modelde madde zorluğu kavramı ortadan kalkmış yerine, yerel eşik (basamak) zorluğu kavramı devreye girmiştir.

---

[37] Tanım gereği $x_j = 0$ için $\kappa_{x_j} = 0$'dır. Yani en düşük şıkkı işaretleyen cevaplayıcının geçtiği toplam eşik sayısı, doğal olarak, sıfırdır.

[38] Bir başka değişle (2.7) denklemi (2.10)'nin özel bir halidir.

[39] Burada $\sum_{h=0}^{0}(\theta - \delta_{jh}) \equiv 0$ olarak kabul edilmektedir. Yani hiç eşik geçemeyen bir cevaplayıcının yoğunluk/zorluk sıralaması açısından ilk sırada olan şıkkı işaretleme olasılığı bire eşittir. Daha güzel bir ifadeyle, kişinin özelliği ne kadar düşük olursa olsun en azından ilk şıkkı işaretleyebilecektir. Bu durumun anlaşılmasında veya kabul edilmesinde tutum ölçümü açısından hiçbir sorun yoktur. Ancak bilgi ve yetenek ölçümü söz konusu olduğunda bu koşul akla yatkın gelmeyebilir.

Örneğin bir ölçekteki ikinci ve beşinci soru zorluk açısından doğrudan birbirleriyle karşılaştırılamayacaktır.[40]

Modeldeki δ_{jh} parametresi 'basamak parametresi' veya 'geçiş parametresi' isimleriyle de anılabilmektedir ve teorik olarak bir sorudaki h şıkkının aynı sorudaki bir önceki şıkka göre nispi zorluğunu (de Ayala, 2009) göstermesi beklenmektedir. Bir başka deyişle bu parametre, o soru çerçevesinde ardışık iki şıktan birisinin seçileceği bilinirken, sonraki olan şıkkın seçilme (koşullu) ihtimalini belirlemektedir. Burada altı çizilmesi gereken husus, aynı sorunun şıkları arasında ölçeği hazırlayanın öngördüğü şekilde bir zorluk sıralaması olacağına dair herhangi bir kısıtın modelin kendisinde bulunmadığı gerçeğidir. Örneğin analiz sonucunda, bir sorunun katılıyorum şıkkını işaretlemek için gereken özellik düzeyinin aynı sorunun tamamen katılıyorum şıkkını işaretlemek için gereken özellik düzeyinden daha yüksek olduğu sonucu ortaya çıkabilir.[41]

Muraki'nin (1992 ve 1997) Genelleştirilmiş Kısmî Puanlama Modeli (GKPM) ise Masters'ın modelinin ayırt edicilik parametresinin dâhil edilmiş halidir.

---

[40] δ_{jh} parametresi, bir şıkkın zorluk açısından tüm test içersindeki konumunu temsil etmemektedir. Basamak parametresinin global (testin bütünü için geçerli olan) eşik zorluk parametresi olarak kabul edilememesinin nedeni, şıkları seçme olasılıkları arasında bağımsızlık koşulunun ihlal edilmesidir.

[41] Bazı âlimler bu durumu ÇİRM'in zayıf noktası olarak görmesine rağmen, Masters'ın parametreleme tercihinden dolayı orta çıkan karışıklık bir kenara bırakılırsa, bu aslında bir sorun olarak görülmemelidir. Şık sıralarının "arzulanan" şekilde olmayabilmesi araştırmacıya ölçeğinin model ile uyumlu olup olmadığını (bir yönüyle geçerliliğini) test etme imkânı vermektedir. Dolayısıyla bu (Andirch'in önerdiği şekilde parametrelendirilirse), modelin üstünlüğü olarak kabul edilmelidir. Ölçek geliştirici, ancak böyle bir durumun (bu duruma bozuk sıralı kategoriler de denilmektedir) ortaya çıkabilmesine olanak verildiği halde şıkların sırası beklentiler doğrultusunda çıkarsa ölçeğinden emin olabilir, beklentilere uymayan bir sıralamanın ortaya çıkmasına izin vermeyen bir modeli temel aldığında değil! Masters'ın yaklaşımının sorunlu olmasının nedeni ise, modelin aslen bilgi ve yetenek ölçümünde kullanılan testler için geliştirilmiş olması ve daha sonra yetenek ölçümü ile tutum ölçümünün aynı kabul edilip, modelin Likert tipi ifadeler için de genellenmesidir.

GKPM şu şekilde ifadelendirebilir:[42,43]

$$P(x_j|\theta, \delta_{jh}, \alpha_j) = \frac{\exp\left[\sum_{h=1}^{k_j} \alpha_j(\theta - \delta_{jh})\right]}{\sum_{c=1}^{m_j} \exp\left[\sum_{h=1}^{c} \alpha_j(\theta - \delta_{jh})\right]} \quad (2.12)$$

Ayırt edicilik parametresin dâhil edilmesinin neticesinde model bir Rasch Modeli olma vasfını kaybeder. Ancak modelin diğer çok ihtimalli madde yanıt modellerinden farklı (Rasch Modellerine daha çok benzeyen) bir işleyiş mekanizması vardır. Dolayısıyla GKPM'nin, Rasch ve iki parametreli çok ihtimalli modellerin ortasında bir konumda olduğu kabul edilebilir. Bu özelliğinden dolayı, bu çalışmada (karşılaştırma yapılan modeller arasında yer almamasına rağmen) veri üretiminde GKPM esas alınmıştır.[44]

2.3.2 Derecelendirilmiş Yanıt Modeli

Çok İhtimalli ölçümlere yönelik olan iki parametreli modellerden en yaygın şekilde kullanılanı Samejina'nın Derecelendirilmiş Yanıt Modelidir[45] (DYM) (1969 ve 1997). Bu modelin altında yatan mantık GKPM'ninkinden oldukça farklıdır ve temeli Thurstone'un kümülâtif sınırlar yaklaşımına dayanmaktadır (Ostini ve Nering, 2006, s. 61). DYM, *m* adet şıktan oluşan çok ihtimalli bir sorunun *m-1* adet iki ihtimalli soru setine dönüştürülebileceği prensibi üzerine kurulmuştur. Örneğin dört şıklı bir tutum ölçeği sorusu; (a) birinci şıkkı seçme veya iki ve üstü şıklardan herhangi birisini seçme, (b) bir ve ikinci şıklardan birisini seçme veya üç ve üstü şıklardan herhangi birisini

---

[42] GKPM burada KPM'nin genelleştirilmiş hali olarak formülize edilmiştir. Muraki (1997) alternatif olarak, Andrich'in formülasyon şeklinin genelleştirilmiş hâline de sunmuştur. (Alternatif gösterim ve açıklamalar için bkz. Ostini ve Nering (2006, s. 50-2).

[43] Burada KPM'den tek farklı olan nokta α_j parametresinin modele dâhil edilmesi değildir. Ayrıca KPM'de model aşılan basamak (x_j) üzerinden tanımlanırken, GKPM seçilen şık (K_j) üzerinden formülleştirilmiştir.

[44] GKPM, bir Rasch Modeli olmamasına rağmen cebirsel açıdan tüm Çok İhtimalli Rasch Modellerini kapsar.

[45] Samejima, benzer mantığa sahip birçok varyasyon geliştirmiştir. Burada en yaygın olarak benimsenen Homojen Durum modeli incelenmektedir.

seçme, (c) bir, iki ve üçüncü şıklardan birisini seçme veya dört ve üstü şıklardan herhangi birisini seçme şeklinde üç adet iki ihtimalli soruya çevrilebilir. Bu yeni oluşturulan sorulardan her birisi iki ihtimalli 2 PL model ile incelenebilir:

$$P^*_{x_j}(\theta) = \frac{\exp\left[\alpha_j(\theta - \delta_j)\right]}{1 + \exp\left[\alpha_j(\theta - \delta_j)\right]} \tag{2.13}$$

Herhangi bir şıkkı işaretleme olasılığı ise;

$$P_{x_j}(\theta) = P^*_{x_j}(\theta) - P^*_{x_{j+1}}(\theta) \tag{2.14}$$

denklemi kullanılarak hesaplanabilir. Tabiî bunun yapılabilmesi için kişinin birinci şık veya üzerini seçme ihtimalinin 1'e, sonuncu (yukarıdaki örnek için dördüncü) şıktan yukarıda bir şıkkı işaretleme ihtimalinin 0'a eşit olması gerekmektedir.

Yukarıda verilen dört şıklı örnek için, mesela kişinin ikinci veya üstü şıkkı seçme ihtimalinden, üçüncü veya dördüncü şıkkı işaretleme ihtimali çıkarılarak o bireyin ikinci şıkkı seçme ihtimali hesaplanmış olur. Şıkların seçilme ihtimallerini hesaplarken yapılan bu işlemlerden dolayı bu model Çok İhtimalli Fark Modelleri ailesine dâhil edilmektedir.[46] Bireyin birinci şıkkı seçme ihtimali ise, iki veya üstünü seçme ihtimalinin 1'den çıkarılmasıyla elde edilir. Son şıkkı seçme ihtimali ise dördüncü şıkkın üstünde bir şık seçme ihtimali (bu örnek için) sıfıra eşit kabul edildiğinden, dört ve üstü şıklardan herhangi birisini seçme ihtimaline eşit kabul edilir.

Bu model çerçevesinde, ayırt edicilik parametresinin testteki sorular arasında farklılık göstermesi mümkün kabul edilirken, aynı soru için geçerli olan şıkların ayırt edicilik parametrelerinin eşit olduğu kabul edilmektedir.

---

[46] Burada bahsedilen isimlendirme Thissen ve Steinberg'in (1986) çokça kabul gören sınıflandırmasına göredir. Bu sınıflandırmaya göre Rasch modelleri ve GKPM ise 'Toplam'a-Böl' türünün örnekleri olarak kabul edilmektedir.

2.4 Parametre Tahminleme Yöntemleri

Tüm Madde Yanıt Modelleri, bilinen kişi ve madde özelliklerinden yola çıkarak kişinin bir ifadeye belli bir şekilde cevap verme olasılığını açıklamaya çalışır. Ancak araştırmacılar çoğu zaman bireylerin ifadeleri nasıl cevaplandıracaklarıyla veya cevapların olasılıklarını tahmin etmekle ilgilenmezler. Aksine, katılımcıların yanıtlarına dair elde kesin bir bilgi vardır. Gerçekte çoğu zaman, eldeki yanıtlardan yola çıkarak, maddenin ve kişinin özellikleri konusunda bilgi sahibi olmak arzulanmaktadır. Dolayısıyla madde ve de özellikle kişi parametrelerini tahminlemek tipik bir ölçme girişiminin asıl amacını oluşturur. Bunu yapabilmek için süreci tersten işletmek gerekmektedir.

Ölçek uygulandıktan sonra denklem (2.7)'de verilen olasılık fonksiyonu, olabilirlik fonksiyonuna dönüşecektir. Ayrıca araştırmacıların elinde kişilerin sadece bir soruya değil, tüm sorulara (L adet soru) verdikleri cevaplar (cevap seti) bulunmaktadır. Dahası araştırmacının elinde sadece bir kişiye ait değil, tüm teste tâbi tutulanlara (N kişi) ait cevap setleri bulunmaktadır. Bir bireyin sorulara verdiği yanıtların diğer sorulara verdiği yanıtlardan bağımsız olduğu kabul edildiğinden ve bireylerin birbirlerinden bağımsız olarak soruları cevaplandırdığı kabul edildiğinden elde bulunan veri matrisinin olabilirlik fonksiyonu (Λ);

$$\Lambda = \frac{\exp\left[\sum_{i=1}^{N}\sum_{j=1}^{L} x_j(\theta_i - \delta_j)\right]}{\prod_{i=1}^{N}\prod_{j=1}^{L}\left[1 + \exp(\theta_i - \delta_j)\right]} \qquad (2.15)$$

şeklinde olacaktır (Wright ve Stone, 1979, s. 62). Bu fonksiyonun türevini sıfır yapan değerler; parametrelerin 'Ençok Olabilirlik Tahminlemeleri' olarak isimlendirilmektedir. Uygulamada hesap kolaylığı sağladığı için Λ fonksiyonunun kendisi değil doğal logaritması ($\ln \Lambda$)[47] esas alınır ve sıfıra eşitlenip denklem çözülmeye çalışılır.

[47] $\ln \Lambda$ yerine genelde λ simgesi kullanılmaktadır. Burada asıl denklemi en çok yapan değer ile bu denklemin logaritmasını en çok yapan değerin aynı olacağı hatırlatılır (ayrıntılar için bkz. de Gruijter ve van der Kamp, 2008, s. 172).

Ancak, madde ve kişi parametreleri bilinirken doğru yanıt verme olasılığını hesaplamakta hiçbir zorluk bulunmazken, sadece verilen yanıtlar bilinirken madde ve kişi parametrelerini tahminlemek, ilave kısıtlamalar veya varsayımlar getirilmedikçe mümkün olamayacaktır. Bu kısıtlamalar ve varsayımlar ile ilgili farklı tercihler, farklı tahminleme yöntemlerinin ortaya çıkmasına neden olmuştur.[48]

'Müşterek Ençok Olabilirlik' (MEO) tahminleme yöntemi kişi ve madde parametrelerinin sıra ile tahminlenmesi prensibine göre işlemektedir. Sorulara verilen yanıtlardan (ham puanlardan) madde parametrelerine ilişkin geçici tahminlerde bulunulur. Bu tahminler gerçek değerlermiş gibi kabul edilerek, kişi parametresi Newton-Raphson 'yenilemeli' prosedürü ile çözülür. Bu işlemler sonucunda kişi parametrelerine ilişkin tahminlemeler elde edilmiş olur. Bu tahminlemeler doğru kabul edilerek madde parametreleri benzer bir süreç işletilerek tahminlenir. Bu elde edilen madde parametreleri, ilk başta yapılan tahminlerden daha iyi olacağı için, madde parametrelerine ilişkin bu yeni tahminlemeler kullanılarak kişi parametreleri yeniden tahminlenir. Her tekrarda tahminlemelerin kalitesi artacağından bu süreç istenilen hassaslığa ulaşana kadar devam ettirilir. Bir önceki tahmin ile bir sonraki tahmin arasındaki ilerleme farkı daha önceden belirlenen bir düzeyin altına düşünce yenileme süreci durdurulur ve elde edilen son tahminlemeler, parametrelere ilişkin nihai tahminlemeler olarak kabul edilir.

Bu yöntem, Kısmî Puanlama Modeli için uygulanırsa, veri matrisinin olabilirlik fonksiyonu (Λ);

$$\Lambda = \frac{\exp\left[\sum_{i=1}^{N}\sum_{j=1}^{L}\sum_{h=0}^{x_{ji}} x_j(\theta_i - \delta_{jh})\right]}{\prod_{i=1}^{N}\prod_{j=1}^{L}\left[\sum_{h=0}^{m_j}\exp\sum_{k=0}^{h}(\theta_i - \delta_{jk})\right]} \qquad (2.16)$$

---

[48] Bu çalışma çerçevesinde yapılan analizler için Winsteps ve Multilog yazılımları kullanıldığından ve bu yazılımlar Müşterek Ençok Olabilirlik [Joint Maximum Likelihood]-Winsteps ve 'Marjinal Ençok Olabilirlik'-Multilog tahminleme yöntemlerini kullandığından burada sadece bu iki yöntemden kısaca bahsedilecektir. Bu ve diğer tahminleme yöntemleri ile ilgili ayrıntılı bilgi edinmek için bkz. Linacre (1999 ve 2004b), Baker ve Kim (2004) ve Hambleton ve Swaminathan (1985).

olacaktır. Parametrelerin tahminlenebilmesi çerçevesinde λ denkleminin (ln Λ) θ_i ve δ_{jh} 'ye göre birinci ve ikinci türevleri bulunur ve Newton-Raphson prosedürü[49] uygulanır (denklemlerin çözüm aşamaları için bkz. Masters ve Wright, 1997, s. 109-10).

Müşterek Ençok Olabilirlik yöntemi ile elde edilen tahminlemelerin yanlı olduğuna[50] yönelik görüşler (açıklama için bkz. Winsteps 2009, s. 412-4), çok parametreli modellerde kullanılabilecek alternatif tahminleme yöntemleri[51] arayışlarını arttırmıştır. Teste katılanların, dağılım özellikleri bilinen bir anakütleden tesadüfî olarak seçildiğinin varsayılması halinde Marjinal Ençok Olabilirlik tahminleme yöntemi (Bock ve Aitkin, 1981) kullanılabilmektedir. Bu yönteme göre madde parametrelerini tahminleyebilmek için kişi parametrelerini bilme zorunluluğu, bireylerin tesadüfî olarak içinden seçildiği kabul edilen anakütlenin olasılık yoğunluk fonksiyonunun kullanılmasıyla ortadan kalkmaktadır:

$$P\left(\underline{x_i}\middle|\underline{w},\underline{v}\right) = \int P\left(\underline{x_i}\middle|\theta_i,\underline{w}\right) g\left(\theta\middle|\underline{v}\right) d\theta \qquad (2.17)$$

---

[49] Newton-Raphson (NR) yöntemine göre denklemin köküne yönelik yapılan bir geçici tahminden, tahmine karşılık gelen Y değerinin denklemin bu noktadaki eğimine oranı çıkarılarak, evvelki tahminden daha iyi bir tahmin elde edilir (bu yeni tahmin, tekrar aynı sürece tâbi tutulursa daha da iyi bir tahmin elde edilir ve bu yenilemelere istenildiği kadar devam edilebilir). Bir başka değişle, x_0 kişi tarafından atanan ilk tahmini, x_1 NR yöntemi ile iyileştirilen bir sonraki tahmini göstermek üzere, $x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$ 'dir. Ancak burada $f(x)$ fonksiyonu λ'nın türevi olduğundan çözüm için λ'nın ikinci türevlerine de ihtiyaç vardır.

[50] Bu yöntemin bir eksikliği de ekstrem kişiler ("sıfır çeken" veya tüm soruları doğru yanıtlayan) ve ifadeler için (kimsenin doğru yapamadığı veya herkesin doğru yaptığı) doğrudan bir tahminleme yapamamasıdır. Bu gibi durumlar için dolaylı çözüm yolları için bkz. de Ayala (2009, s. 59)

[51] Koşullu Ençok Olabilirlik gibi bazı diğer alternatif yöntemler sadece Rasch Modellerinde kullanılabilmektedir.

ki burada, $\underline{x_i}$ bir kişiye ait yanıtları içeren vektörü, $g(\theta|\underline{\upsilon})$ ise katılımcılar anakütlesindeki yeteneklerin olasılık yoğunluk fonksiyonunu[52] temsil etmektedir. $P\left(\underline{x_i}\middle|\theta_i, \underline{\psi}\right)$ ise madde parametreleri[53] ve kişi özelliği bilinirken bireye ait cevap setinin olasılığını göstermektedir. Yukarıdaki denklem bir kişi için yanıt seti olasılığını açıklarken, ölçüme dâhil olan herkesten elde edilen sonuçlarla ilgili olabilirlik fonksiyonu;

$$\Lambda = \prod_{i=1}^{N} \int P\left(\underline{x_i}\middle|\theta_i, \underline{\psi}\right) g(\theta|\underline{\upsilon})\, d\theta \tag{2.18}$$

olacaktır. Böylece kişilerin yetenek parametreleri bilinmediği halde, yani kişi parametreleri doğrudan tahminlemek zorunda kalınmadan (sadece kişilerin nasıl bir anakütleden[54] geldiği bilinerek) madde parametreleri tahminlenebilmektedir.[55] Madde parametrelerinin tahminlenmesinin, kişi parametrelerinin tahminlenmesinden bağımsız olmasının bedeli, parametre tahminlerinin kişilerin dağılım şekline bağımlı hale gelmesidir.[56]

---

[52] $\underline{\upsilon}$ burada dağılım parametrelerini içeren vektörü göstermektedir. Normal dağılım için bu parametreler ortalama ve varyans iken mesela dikdörtgensel dağılım için bu parametreleri alt ve üst limitler oluşturur.

[53] $\underline{\psi}$ madde parametrelerinden oluşan matrisi temsil etmektedir.

[54] Anakütlenin genelde normal dağıldığı varsayılır. Böyle olması bir koşul olmamasına karşın hemen hemen tüm beşeri özelliklerin dağılımına ilişkin en makul tahminin normal dağılım olması nedeni ile nerdeyse tüm ölçümlerde anakütlenin standart normal dağıldığı kabul edilir. Burada genelde kabul edilen ilave bir kısıt ise tüm katılımcıların aynı anakütleden geldiğinin varsayılmasıdır.

[55] Uygulamada yukarıdaki denklemin çözülmesi yerine Hermite-Gauss Kareleme [*quadrature*] yöntemi uygulanarak eğrinin altındaki alan yaklaşık olarak hesaplanır. Yöntemin Çok İhtimalli Modeller için uygulanışı da aynı şekilde olmaktadır.

[56] Bu madde yanıt kuramının değil, tahminleme yönteminin bir eksikliğidir.

Bir kez madde parametreleri tahminlendikten sonra, klasik Ençok Olabilirlik yöntemi ile veya Bayesçi yöntemler kullanılarak kişi parametreleri tahminlenebilir (de Ayala, 2009, s.68-77).

2.5 Madde Yanıt Kuramının Genel Varsayımları

Madde Yanıt Kuramı çerçevesindeki modellerin en önemli varsayımı ölçekteki tüm ifadelerin aynı özelliği ölçtüğünün kabul edilmesi anlamına genel tekboyutluluk varsayımıdır.[57] Bir başka deyişle, bu varsayıma göre kişilerin ifadelere belirli bir şekilde yanıt verme olasılıkları, bireylerin sadece bir (ve herkes için aynı olan) sürekli örtük özelliğinin fonksiyonu olmalıdır (de Ayala, 2009, s.20). Örneğin, aritmetik işlem yapabilme yeteneğini ölçen bir test sorusunu doğru yanıtlama olasılığı sadece, kişinin bu konudaki yeteneğine bağlı olmalıdır. Kişinin soruyu doğru yanıtlayabilmesi için, mesela üç boyutlu düşünme yeteneğinin yüksek olması, el becerilerin iyi olması veya integral almayı bilmesi gerekmemelidir.

Tekboyutluluk varsayımının sosyodavranışsal araştırmalarda mükemmel bir şekilde tutturulması hemen hiçbir zaman mümkün değildir. Bunun nedeni, beşeri özelliklerin (davranışların, yeteneklerin, inançların v.b.) karmaşık yapılar olmaları ve genelde birden çok bileşen içeren özellikte olmalarıdır[58] (Andrich, 1988, s. 9-10). Ancak tekboyutluluk zaten ideal olan durumu tanımlamak için kullanılan bir kavramdır ve mutlak değil nispi olarak ele alınmalıdır. Dolayısıyla bir düzeye kadar bu varsayımın ihlal edilmesi, sonuçların işe yararlılığını ciddi bir şekilde düşürmeyecektir.

MYK modellerinin ikinci bir varsayımı koşullu bağımsızlık (veya yerel bağımsızlık) olarak isimlendirilmektedir. Bu kavram ile teste tâbi olan bir kişinin bir

---

[57] Zamanla çokboyutlu modeller de geliştirilmiştir. Bu modellere göre bu kısıt, "tüm maddelerin sadece belirli bazı özellikleri aynı anda ölçtüğü" şeklinde gevşetilmiştir. Ancak çokboyutlu modellerin uygulanış alanları oldukça sınırlıdır. Yalnız burada, MYK bakış açısıyla ele alınan çokboyutluluk olgusu ile örtük değişken-faktör analizi çerçevesinde ele alınan çokfaktörlülük kavramlarının ayını olmadığını hatırlatmak gerekir.

[58] Mesela, herhangi bir kâğıt-kalem testinde (test neyi ölçerse ölçsün) bir yeteneğin izole edilip ölçülebilmesi oldukça güçtür. Bireyin, ölçülen özelliğinin dışında en azından okuma-yazma yeteneğinin olması gerekir ve bu konudaki yeteneği sonuçları etkileyecektir.

soruya vereceği yanıtın, diğer sorulara verdiği yanıtlardan etkilenmemesi koşulu kastedilmektedir. Yani, soruyu doğru olarak yanıtlama ihtimalini, sahip olunan örtük özellik düzeyinden başka bir şeyin (örneğin önceki soruları doğru yanıtlama ihtimalinin) etkilememesi gerekmektedir. Bu açıdan ele alındığında koşullu bağımsızlık varsayımının, tekboyutluluk varsayımı ile aynı (ya da bu varsayımın bir parçasını oluşturduğu) düşünülebilir. Gerçekten de tekboyutluluk varsayımının mutlak bir şekilde sağlanması durumunda, koşullu bağımsızlık varsayımı da bizatihi sağlanmış olur. Ancak bunun tersi doğru olmadığı için ve bu varsayımın tahminleme yöntemleri açısından kritik bir önemi olduğu için hemen hemen tüm kaynaklarda iki varsayım ayrı ayrı ele alınmıştır.

Bu varsayımın ihlal edileceği durumlara örnek olarak; bir sorudan elde edilen cevabın diğer sorunun çözümünde mutlaka kullanılmak zorunda olunması, bir okuma paragrafı ile ilgili birden fazla sorunun bulunması[59] veya bir soruda, testteki başka bir soruya ilişkin bir ipucunun bulunması gibi durumlar verilebilir.[60]

Koşullu bağımsızlık varsayımının istatistikî olarak test edilmesinden çok, ölçüm aracı geliştirilirken bunun dikkate alınması ve mantıksal olarak bu varsayıma ters düşen bir ölçek tasarımının yapılmaması önemlidir.[61,62]

---

[59] Her okuma paragrafına ait sadece bir sorunun sorulması pek ekonomik olmadığından, birden fazla soru sorma uygulaması benimsenip o paragrafa ait tüm soruların, çok ihtimalli "süper soru" olarak ele alınması ara çözüm olarak kullanılabilmektedir (örnek için bkz. Baghaei. 2008).

[60] Psikometri literatüründe, yerel bağımsızlık olgusunun bu yönü sadece önceki soruların, bir sonraki soruyu doğru yanıtlama ihtimalini artırması ile sınırlı tutulmuştur. Hâlbuki bir soruyu doğru yanıtlamanın diğer soruyu yanış yanıtlama ihtimalini artırabilmesi de söz konusu olabilir. Buna benzer durumlar için psikoloji bilimindeki 'Einstellung etkisi' veya 'problem çözmede mekanizasyon' olguları geliştirilmesine karşın bu ihtimal ölçme kuramcılarının ilgisini çekmemiştir. Söz konusu kavramlar ile ilgili olarak bkz. Luchins ve Luchins (1950).

[61] Hableton ve Swaminathan (1985), "yerel bağımsızlığın; bir deneğin (tüm teste yönelik) cevap setinin olasılığının, o deneğin her bir ifadeye verdiği yanıtların ayrı ayrı olasılıklarının çarpımına eşit olması (ve bunun tüm denekler için doğru olması) durumunda

3. ARAŞTIRMA YÖNTEM BİLİMİ VE İZLENEN PROSEDÜRLER

Bu bölümde araştırmanın çeşitli özellikleri, yöntem bilimi açısından incelenmiştir. Önce araştırmanın genel tasarım özelliklerinden bahsedilmiş, daha sonra araştırmanın değişkenleri, maddî hipotezleri, sınırlılıkları tanıtılmıştır. Simülasyon yöntemi ile ilgili kısaca bilgi verildikten sonra başvurulan veri analiz yönteminin bazı özellikleri açıklanmıştır.

3.1 Araştırmanın Tasarımına Genel Bakış

Madde Yanıt Kuramı çerçevesinde geliştirilen modeller ile ilgi yapılan simülasyon araştırmalarında çok farklı amaçlar güdülebilmektedir. Ancak çalışmaların büyük bir kısmı bir modelin farklı koşullar altındaki performansını incelemek veya ilgilenilen bazı modellerin, belirlenen koşullar altındaki performanslarını karşılaştırmak amacıyla yapılmıştır. Modellerin hangi konudaki performanslarının inceleneceği çalışmadan çalışmaya farklılıklar gösterebilmektedir. Ancak modeller en çok model-veri uyumu ve madde parametrelerini tahminlemedeki performansları açısından mukayese edilmekte veya kendi içlerinde incelenmektedir. Model-veri uyumunun incelendiği simülasyon çalışmalarında hangi madde yanıt modelinin eldeki verileri daha iyi açıkladığı[63] veya belli bir modelin mevcut verileri hangi koşullarda daha iyi açıkladığı araştırılmaktadır.

İlginç bir şekilde modellerin kişi parametresini tahminlemedeki performanslarını inceleyen çalışmaların sayısı oldukça azdır. Neticede herhangi bir

---

geçerli olacağını" belirtmiştir (s. 23). Ancak konunun bu şekilde ele alınması tarafımca, "nedenler ile sonuçların birbirine karıştırılması" olarak yorumlanmaktadır.

[62] Bu varsayımların dışında; (a) ölçüm aracının hız testi değil de, güç testi olması gerektiği (Hableton ve Swaminathan, 1985) ve (b) verilerin modelce belirlenen fonksiyona uyması gibi (de Ayala, 2009) iki varsayımdan daha bahsedilmektedir. Ancak bunlardan ilki tekboyutluluk varsayımının uygulamaya yönelik bir yönünden ibaret olduğu için ve ikinci unsurun aslında bir varsayımı değil modelin kendisini tasvir ettiği düşünüldüğünden burada ayrıca incelenmeyecektir.

[63] Daha öncede belirtildiği gibi tarafımızca, bunun sağlıksız bir yaklaşım olduğu düşünülmektedir. Bu çalışmada ölçme girişimcisinin amacının elindeki "kutsal" kabul ettiği verileri en iyi açıklayan modeli tespit etmek değil, doğru kabul ettiği modele uyan veriler elde etmek olması gerektiğini düşüncesi savunulmaktadır.

bilimsel ölçme girişiminin temel amacının nesnelerin ilgilenilen özelliklerini mümkün olduğu kadar doğru ve hassas bir şekilde tahmin etmek olduğu düşünülürse bu konuda bu kadar az çalışma olması oldukça gariptir. Araştırmacıların, model-veri uyumu veya madde parametrelerinin tahminlenmesindeki başarıyla ilgilenmeleri normaldir çünkü bu kriterlerin nihayetinde kişi özelliğini tahminleme sürecini etkilediği bilinmektedir.[64] Dolayısıyla bu tip araştırmalara da ihtiyaç vardır. Ancak asıl olarak modellerin farklı koşullar altında kişi parametrelerinin tahminlenmesindeki performanslarının yoğun bir şekilde incelenmesine ihtiyaç vardır. Bu nedenle bu çalışmada doğrudan kişi parametrelerinin tahminleme kalitesi kriter olarak alınmıştır.

3.2 Araştırmanın Bağımlı Değişkenleri ve Araştırmanın Analiz Birimi

Ölçmenin tanımı gereği, bir ölçme girişiminde bulunulduğunda nesnelerin kendilerinin değil, belirli bazı özelliklerinin (atıflarının) ölçülmesinden bahsedilebilir. Örneğin, teknik açıdan doğru olma kaygısı taşınıyorsa, bir çalışmada öğrencilerin değil, öğrencilerin sayısal yeteneklerinin ölçüldüğünden bahsedilebilir. Nesneler, sayısız açıdan ölçülebilir veya birbirleriyle mukayese edilebilir. Nesnelerin hangi özellikleriyle ilgilenildiği, dolayısıyla hangi özelliklerinin ölçüleceği araştırmanın değişkenlerini oluşturacaktır. Nesnelerin (varlıkların) türü veya seviyesi ise araştırmanın analiz birimini oluşturur (bkz. Trochim, 2006). Sosyodavranışsal araştırmalarda analiz birimi genelde bireyler, gruplar, örgütler veya kültürlerdir. Araştırma sorusu ve konusu, analiz biriminin örneğin, birey mi, yoksa örgüt mü olacağını belirlerken, analiz birimi ise, araştırmada başvurulacak örnekleme yöntemini, olguların sayısallaştırılış şekillerini (değişkene dönüştürülüş sürecini) ve verilerin incelenmesinde kullanılacak istatistiksel analiz yöntemlerini belirleyici bir konumda olacaktır. Madde yanıt kuramı ile ilgili

---

[64] Örneğin, madde zorluk düzeylerini mümkün olduğu kadar doğru bir şekilde tahminlemek, bireylerin yetenek düzeylerinin de daha isabetli ve yansız bir şekilde tahminlenmesini sağlayacaktır. Dolayısıyla yetenek düzeylerinin tahmin kalitesinin temel amaç olmadığı istisnai durumlar dışında (mesela soru bankası oluşturabilmek için madde parametrelerini tespit ederken) bahsettiğimiz diğer kriterler bir nevi ara hedef işlevi görmektedir.

yapılan simülasyon çalışmalarının birçoğunda analiz birimi olarak kalibrasyon grubu kullanılırken bu çalışmada analiz birimi bireysel[65] düzeyde ele alınmıştır[66].

"Neden" sorusuna[67] cevap verme ihtiyacı bilimin nüvesini oluşturduğu için, bir bilimsel araştırmanın temelini de doğal olarak nedenlerini merak ettiğimiz olgu oluştur. Bu nedenlerini[68] (etkileyenlerini, öncüllerini) merak ettiğimiz olgu genelde araştırmanın bağımlı değişkeni[69] olarak isimlendirilmektedir. Araştırmacılar en çok bu değişkenlerle ilgilenirler (Thompson, 2006). "Bağımlı değişken, bilimsel bir araştırmanın temelini oluşturur" ifadesiyle şu kastedilmektedir: Hemen hemen tüm

---

[65] Bu bir simülasyon çalışması olduğu için burada gerçek bireylerden değil sanal bireylerden [*simulee*] bahsedilmektedir.

[66] Kenny (2003) değişmez ve kesin bir kural olmadığını belirtmekle birlikte bir araştırmada analiz birimi ile atama biriminin aynı olmasını önermektedir. Bu çalışmada analiz birimi ile atama birimi arasında bir uyumsuzluk varmış gibi durmaktadır. Ancak ileride belirtileceği üzere çalışmada *bütünüyle birimler-içi tasarım* kullanıldığı için atama biriminden bahsedilemez. Analiz biriminin birey düzeyinde tutulmasının nedeni ise yineleme sayısını azaltmaktır.

[67] Günlük hayatta kullanıldığı haliyle neden sorusunun iki farklı anlamı (amaç ve mekanizma olmak üzere) ve bilimin sınırlılıkları ile ilgili olarak Helen Quinn'in (2009) kısa ama bilgilendirici köşe yazısına bakılabilir. Nedensellik olgusuyla ilgili olarak ayrıca bkz. Punch (2005, s. 48-52).

[68] Olguların nedenlerini araştırmak bilimin temel amacı olmasına karşın "neden" kelimesini [*cause*] bilimsel araştırmalarda kullanmaktan özellikle kaçınılır. Bilim adamı, bu kelimeyi kullanmak zorunda kalmış ise genelde dipnotlarla, açıklamalarla adeta özür diler veya neden kelimesini teknik anlamıyla değil günlük anlamıyla kullandığını belirtir. Neden kelimesini sadece bilim felsefecileri serbestçe kullanabilir!

[69] "Bağımlı değişken" ifadesinin sadece deneysel çalışmalarda kullanılmasına ilişkin tavsiyeler olmasına karşın araştırmacalar tarafından bu tavsiyelere pek riayet edilmez (Licht, 1995). Kriter değişkeni [*criterion*], sonuç değişkeni [*outcome variable*], tepki değişkeni [*response variable*], içsel değişken [*endogenous variable*] gibi kavramlar da eş veya benzer anlamlı olarak kullanılabilmektedir. Aynı olgu için bu kadar farklı kavramın kullanılması, Thompson'un (2006, s. 4) "istatistikte bir olgu ne kadar önemli ise, o kadar çok farklı eş anlamlısı kullanılır ki öğrencilerin kafası daha çok karışsın" nükteli hipotezini kanıtlar niteliktedir.

bilimsel girişimlerin çıkış noktası, "A olgusu acaba hangi olguları etkiliyor?" tipi sorulara değil "B olgusu hangi olgulardan etkileniyor (B olgusunun nedenleri nelerdir)?" tarzı sorulara cevap bulmaktır. Ancak bu, ilk soru tipinin önemsiz, bu soruya cevap bulmaya çalışmanın da gereksiz olduğu anlamına gelmemektedir. Ancak tipik bir bilimsel çalışmanın varlık nedeni, bağımlı değişken ile ilgili olarak araştırmacının duyduğu meraktır. Diğer değişkenler sadece, bağımlı değişkene ilişkin merakın giderilmesinde kullanılan araçlardan ibarettir.

Bu çerçevede düşünüldüğünde, bu çalışmanın temelini kişilerin özelliklerine ilişkin yapılan tahminlemelerin performansını etkileyen unsurları belirleme arzusunun oluşturduğu söylenebilir. Başka bir deyişle, ('olgu' düzeyinde[70] ele alacak olunursa) nedenleri merak edilen kavram tahminlemelerde ne derece başarılı olunduğudur.

Tahminlemelerin performansı farklı açılardan incelenebilir. 'Yapı' düzeyinde bahsedecek olunursa, bu çalışmada performans, tahminlemelerdeki isabetlilik ve istikrar olmak üzere iki açıdan incelemiştir. Bu yapıların, 'değişken' olarak araştırma modeline dâhil edilebilmesi için amelî tanımlarının[71] oluşturulması (olguların amelîleştirilmesi) gerekmektedir.

Tahminlemelerdeki 'isabetlilik' farklı şekillerde amelîleştirilebilir. Mesela en basit haliyle tahminleme değerinden gerçek özellik değerinin çıkarılması şeklinde hesaplanabilir. Ancak bu çalışmada isabetlilik;

$$\left(\bar{\hat{\theta}}_i - \theta_i\right)^2$$

şeklinde tanımlanmıştır ki burada θ_i gerçek yetenek düzeyini, $\bar{\hat{\theta}}_i$ ise tahminlemelerin yinelemeler boyu ortalamasını temsil etmektedir. Bu şekliyle kullanıldığında, isabetlilik kavramı yerine daha çok 'tahminleme yanlılığı' kavramı kullanılmaktadır.

Tahminlemelerdeki 'istikrar' kavramıyla ise aynı gerçek yetenek düzeyine ilişkin yinelemeler boyunca elde edilen tahminlemelerin ne kadar değişkenlik gösterdiği kastedilmektedir. Bu çalışmada istikrar;

---

[70] Aynı kavramdan, olgu, yapı ve değişken olmak üzere üç farklı düzeyde bahsedilebilir. Ayrıntılar için bkz. Kerlinger ve Lee (2000, s. 42-43).

[71] Araştırma yöntem bilimi açısından çok önemli olan İngilizcedeki '*operational definition*' kavramı yerine, 'operasyonel', 'işlemsel' veya 'işevuruk' tanım kavramları da kullanılabilmektedir.

$$\left(\sum_{r=1}^{R}\left(\hat{\theta}_{i_r}-\bar{\hat{\theta}}_i\right)^2\right)/R$$

şeklinde amelîleştirilmiştir ki burada R yineleme sayısını temsil etmektedir. Bu şekliyle ele alındığında, istikrar kavramı yerine daha çok 'verimlilik' veya 'görgül hata varyansı' kavramları kullanılmaktadır.

Bu çalışmada bağımlı değişken olarak modele dâhil edilen tahminleme yanlılığı ve görgül hata varyansı olgularının dışında benzer çalışmalarda farklı bağımlı değişkenler de kullanılabilmektedir. Mesela Farkların Karelerinin Ortalaması (FKO) kriter olarak alınabilecek değişkenlerden birisidir ki aslında cebirsel olarak tahminleme yanlılığı ve görgül hata varyansının toplamına eşittir:

$$\left[\sum_{r=1}^{R}\left(\hat{\theta}_{i_r}-\theta_i\right)^2\right]/R=\left(\bar{\hat{\theta}}_i-\theta_i\right)^2+\left[\sum_{r=1}^{R}\left(\hat{\theta}_{i_r}-\bar{\hat{\theta}}_i\right)^2\right]/R \qquad (3.1)$$

Ayrıca analiz birimi olarak kalibrasyon grubunun[72] seçilmesi halinde birey düzeyli araştırmalarda kullanılamayan bazı değişkenler modele dâhil edilebilir. Mesela bu tip çalışmalarda kriter değişkeni olarak Fark Karelerinin Ortalamasının Karekökü (FKOK);

$$FKOK=\sqrt{\frac{\sum_{i=1}^{N}\left(\hat{\theta}_i-\theta_i\right)^2}{N}} \qquad (3.2)$$

ki burada N kalibrasyon grubunun büyüklüğünü temsil etmektedir, kullanılabilir veya gerçek değerler ile tahminlenen değerler arasındaki korelasyon $\left(r_{\hat{\theta}\theta}\right)$ incelenebilir.

[72] Bu tür çalışmalarda analiz birimi kalibrasyonu grubu olmasına rağmen, ölçüm birimi bireylerdir. Analiz birimi, ölçüm birimi ve atama birimi kavramlarıyla ilgili bkz. Kenny (2003).

3.3 Araştırmanın Bağımsız Değişkenleri

Çalışmada manipüle edilen dört değişken yer almaktadır. Bunlar; (a) tahminlemede kullanılan ölçme modeli, (b) kalibrasyon örneklemi (veya sadece örneklem[73]) büyüklüğü, (c) ölçek uzunluğu (veya ifade sayısı) ve (d) şık/kategori sayısıdır.

Araştırmadaki temel bağımsız değişken tahminlemede kullanılan ölçme modelidir. Araştırmanın çıkış noktası ve birincil amacı bu değişkenin bağımlı değişken üzerinde bir etkisinin olup olmadığını incelemektir. Ancak tipik bir araştırmada olduğunun aksine burada iddia edilen, değişkenin bir düzeyinin (yani bir modelin) diğerinden bağımlı değişken açısından farklı olduğu değil, aralarında fark olmadığıdır[74]. Çalışmada iki model karşılaştırılmaktadır. Bir başka deyişle bu bağımsız değişken iki düzeyden oluşmaktadır. Karşılaştırılan modeller (a_1) Çok İhtimalli Rasch Modeli (Andrich, 1978 ve 2008, Masters, 1982 ve Luo 2005) ve (a_2) Derecelendirilmiş Yanıt Modelidir (Samejima, 1997). Gerçekte bu modeller araştırmamız için birer araç işlevi görmektedir. Aslında bu modeller üzerinden genel olarak Rasch Modeli (veya Tek Parametreli Lojistik Model) ve İki Parametreli Madde Yanıt modelinin çok ihtimalli versiyonları karşılaştırılmaktadır.

Çalışmada, model faktörünün bağımlı değişkenler üzerinde önemli bir etkisi olmadığının kanıtlanmaya çalışılmasının yanı sıra kalibrasyon grubunun büyüklüğünün, (yani test/ölçüm seansına kaç kişinin katıldığının) bağımlı değişkenler üzerindeki olası etkisi de incelenmiştir. Bu değişken de iki düzeyde ([b_1] 150 kişi ve [b_2] 300 kişi olmak üzere) incelenmiştir: Bu düzeyler örgütsel davranış araştırmalarında genelde kullanılan örneklem büyüklükleri dikkate alınarak belirlenmiştir. Eğitim bilimlerinde yapılan araştırmalardaki örneklemlerin çok daha

---

[73] Genelde kalibrasyon örneklemi kavramı yerine kısaca örneklem ifadesi de kullanılabilmektedir. Ancak çalışma istatistiksel analiz içeriyorsa bu bir karışıklığa neden olabilir. Çünkü araştırmanın örneklemi ile kalibrasyon örneklemi farklı büyüklüklerde olabilir. Nitekim bu çalışmada örneklem büyüklüğü 150'ye sabitlenmiş iken, kalibrasyon örneklemi büyüklüğü, ölçüm seanslarının yarısı için 150, diğer yarısı için ise 300 olarak belirlenmiştir.

[74] Farksızlık hipotezinin reddedilememesinin arzulanmasına Kabul-Destek (Accept-Support [AS]) durumu denilmektedir. Kabul-Destek ve Ret-Destek araştırmaları arasındaki farklılıklar için bkz. Steiger ve Fouladi (1997).

büyük olmasına karşın örgütsel davranış araştırmalarında kullanılan örneklem büyüklükleri çok daha sınırlıdır[75].

Çalışmadaki üçüncü ve dördüncü bağımsız değişkenler, ölçeğin özellikleri ile ilgilidir. Ölçekte kullanılan ifade sayısının artırılmasının, diğer her şey aynı kalmak koşulu ile, ölçümün daha az hata ile yapılmasını sağlayacağı bilinmektedir[76]. Ancak "diğer her şey" gerçekte hiçbir zaman aynı kalmamaktadır. En basitinden, soru sayısının artması (araştırmada en az birkaç olgunun daha olduğu ve bunların ölçümü için ayrı ölçekler olduğu da düşünülürse) cevaplayıcıların araştırmaya katılma isteklerini, dikkatlerini ve samimiyetlerini olumsuz yönde etkileyecektir. Bu yüzden genelde ölçüm hassaslığından biraz fedakârlık edip mümkün olan en kısa ölçeklerle ölçüm yapılmaya çalışılmaktadır. Ancak "yeterli hassaslıkta bir ölçüm için gereken en az ifade sayısı nedir?" "soru sayısı ve tahminleme kalitesi arasındaki dengeyi araştırmacı nasıl kurmalıdır?" gibi sorulara yanıt aramak amacıyla ölçek uzunluğu 6, 9 ve 12 soru olmak üzere üç farklı düzeyde incelenmiştir. Yine bu düzeyler tipik olarak örgütsel davranış araştırmalarında bir boyut için kullanılan ifade sayıları dikkate alınarak belirlenmiştir.

Aynı şekilde şık sayısı da ölçümdeki hassaslığı etkilemektedir. Teorik olarak şık sayısı sonsuza yaklaştıkça (kesikli veri türünden sürekli veri türüne yaklaşılmasıyla), sonuçlar daha az hatalı olacaktır. Buna rağmen mesela 100'lü Likert tipi ifadelerin kullanılmamasının nedeni insanın bir konudaki düşüncesini (veya onu temsil eden hayalî uzanımı) yoğunluk bakımından ardışık 100 eşit (veya sıralı) parçaya bölemeyeceği geçeğidir. Genelde tutum ve benzeri olguların ölçümünde (sıklık sırasına göre en çok) 5'li, 4'lü, 7'li, 3'lü ve 2'li ifadeler kullanılmaktadır.

---

[75] İki disiplin arasındaki bu farkın en önemli nedeni eğitim bilimlerinde bireylere (çoğunlukla öğrencilere) ulaşmanın maliyetinin çok düşük olması, buna karşın örgütsel davranış araştırmalarında bireylere (çoğu zaman çalışanlara ve özellikle yöneticilere) ulaşmanın ve onları anket doldurmak için vakit ayırmaya ikna etmenin çok zor olmasıdır.

[76] Paralel formlar ilave edilerek testin uzatılmasının güvenirliği arttırması Klasik Test Kuramının aksiyomatik özelliğidir. Bu prensip 'Spearman-Brown Kehaneti' olarak isimlendirilir ve; $\rho_{CC'} = k\rho_{tt'}/[1 + (k-1)\rho_{tt'}]$ şeklinde formülleştirilir (ayrıntılar için bkz. Crocker ve Algina, 2006) ki burada $\rho_{CC'}$ uzatılmış testin güvenilirliğini, k testin, kendisinin kaç katı kadar uzatıldığını ve $\rho_{tt'}$ ise orijinal testin güvenirliliğini göstermektedir.

Yukarıda bahsedilen unsurlar çerçevesinde şık sayısı değişkeni, 2, 4 ve 6 şık olmak üzere üç düzeyde incelenmiştir. Burada iki hususa dikkat çekmek gerekmektedir. Birincisi, 2'li ifadeler, taban (baz) oluşturması açısından araştırmaya dâhil edilmiştir. Bu ifadelerin yer aldığı benzetimlerin tahminlenmesinde Çok İhtimalli Rasch Modeli, Basit Rasch modeline; Derecelendirilmiş Yanıt Modeli ise İki Parametreli Lojistik Modele dönüşmektedir. İkincisi, uygulamada en sık kullanılan 5'li Likert ölçeğinin düzeyler arasında yer almamasının nedeni 'polinom eğilimi (diğer adıyla, çokterimli eğilim) planlı mukayesesini' yapabilmek için düzeylerin aralıklarının eşit olması gerektiği zorunluluğudur (Green, Salkind ve Akey, 1997 s.215). Alternatif olarak 3, 5 ve 7'li Likert ifadeleri de düzey olarak seçilebilecekken bu durumda taban olarak alacağımız 2'li ifadeler araştırma dışında kalacağı için[77] ve orta şıkla ilgili sorunların gündeme getirilmesi istenildiği için bu alternatif tercih edilmemiştir.

Orta şık sorunu ile şu kastedilmektedir: İster klasik ister modern test kuramı çerçevesinde hazırlansın bir ölçekte tek sayıdan oluşan şıkların kullanılması şıkların cevaplayıcılar arasında farklı yorumlanmasına neden olabilmektedir. Bazı cevaplayıcılar, (test hazırlayıcının da umduğu gibi) orta şıkkı zihninde oluşturduğu hayali uzanımın ortası olarak yorumlarken, bazı cevaplayıcılar bu şıkkı, soruyla ilgili yeterli bilgisi olmadığı, soruyu anlamadığı veya ifade onu ilgilendirmediği durumlarda işaretlenecek bir şık olarak yorumlamaktadırlar. Özellikle bu şık için etiket olarak "kararsızım" veya "ne katılıyorum ne katılmıyorum" gibi ifadeler kullanıldığında veya bu şık sadece soru işareti (?) ile temsil edildiğinde bu farklı anlaşılmalar daha da artabilmektedir. Bu durumu engellemek için genelde 5 yerine 4 şıklı ifadelerin kullanılması önerilmektedir. Bu tarz bir şıklandırma kullanılan ifadelere 'zorlayıcı şıklı ifadeler' de denilmektedir ve aslında bu ifadeler tutum teorisi ile daha uyumlu bir yapıdadır. Bir kere oluştuktan sonra artık tutumların nötr olup olamayacağı sosyal psikoloji literatüründen tartışıla gelen bir meseledir ancak ağırlık; tutum objesi ile ilgili, kişinin bir tutum oluşacak kadar tecrübesi varsa artık o objeye karşı kişinin nötr

[77] Tüm bu niyetlere rağmen, (veri analizi bölümünde açıklanacağı üzere) iki ihtimalli ifadeler daha sonra genel analizden çıkarılmıştır.

(tarafsız) olamayacağı görüşü yönündedir[78] (ayrıntılar için bkz. Krech, Crutchfield ve Ballachey, 1971, Birinci Cilt s. 258-9; ayrıca Baysal, 1981, s. 66-7). "Bir tutum objesine ilişkin kişinin tutum oluşacak kadar bilgisi, tecrübesi varsa kişi çok az da olsa lehte veya aleyhte bir pozisyona sahip olacaktır" şeklindeki görüşün daha çok kabul görmesine karşın tutumların ölçülmesinde zorlayıcı şıklar pek kullanılmamaktadır. Bu konuda araştırmacıları orta şıksız (sıfır noktasız) ifadeler kullanmaya yönlendirmek amacıyla araştırmada 2, 4 ve 6 şıklı ifadeler düzey olarak ele alınmıştır.

Tüm bağımsız değişkenlerin düzeyleri iradî olarak belirlendiği için bu çalışma 'sabit etkiler' araştırma türündendir ve 2x2x3x3 şeklinde bir tasarıma sahiptir.

3.4 Araştırmanın Kontrol Değişkenleri ve Araştırmanın Sabitleri

Genelde bir araştırmada belirlenen bağımsız değişkenlerin dışında bağımlı değişkenin üzerinde veya bağımsız–bağımlı değişken bağlantısının üzerinde etkisi olduğundan şüphelenilen (Miller ve Whicker, 1999, s.27-30) ancak araştırmacının incelemek istemediği veya bağımsız değişken olarak modeline dâhil etmek istemediği değişkenler söz konusu olmaktadır (Punch, 2005, s.67-68). Araştırmacının ilgilenmek istememesi, bu değişkenlerin olası etkilerinin görmezlikten gelinebileceği anlamına gelmemektedir. Araştırmacı sonuçların gerçek ilişkileri yansıtabilmesini sağlamak adına bu değişkenlerin etkilerini denetim altına almak zorundadır. Etkileri denetlenen bu değişkenlere kontrol değişkenleri denilmektedir. Bu etkiler farklı şekillerde kontrol edilebilmektedir. Bu yöntemlerden bir tanesi, kontrol değişkeninin tek bir değer alması sağlanarak o değişkeni araştırmanın sabiti haline dönüştürmektir ki bu çalışmada bu yöntem kullanılmıştır.[79]

---

[78] Sıfır noktalı ifadelerin yarattığı sorunlarla ilgili yapılan araştırma sonuçları için ayrıca bkz. Stone (2004), DuBois ve Burns (1975) ve Kulas, Stachowski ve Haynes (2008).

[79] Sıkça kullanılan diğer iki denetim altına alma stratejisi; (a) kontrol değişkenin değerlerinin gruplar arasında eşit (veya en azından tesadüfî) dağılmasını sağlamak ve (b) istatistiksel kontrol yöntemlerini (genelde kovaryans analizini) kullanmaktır. Kontrol değişkenini araştırmanın sabiti haline dönüştürmenin genellenebilirliği daraltması gibi bir sakıncası olmasına karşın bu çalışma için en uygun yöntem olduğu düşünülmüştür.

Araştırmanın sabitleri arasında, örtük özelliklerin dağılımının şekli[80] [$\sim N(0;1)$], ifadelerin zorluk düzeylerinin dağılımının şekli [$\sim N(0;1)$], ayırt edicilik parametresinin dağılımının şekli[81] [$\sim U(0,8;2)$] gibi unsurlar yer almaktadır. Bir başka deyişle parametrelerin dağılım şekillerinin (mesela aşırı çarpık olması gibi) kişi parametresinin tahminlenmesi üzerinde bir etkisi olabileceği bilindiği halde araştırma tasarımını sadeleştirmek adına bu unsurlar sabitlenmiştir.[82]

3.5 Araştırmanın Maddî Hipotezleri

Bilimsel araştırmaların iskeletini araştırma soruları oluşturur. Araştırma soruları, daha ayrıntılı, açık ve sınanabilir iddialar haline getirilmesiyle maddî hipotezler oluşturulmuş olur. Bu hipotezler, araştırmacının araştırma sorularına ilişkin tahminlerini yansıtır. Bilimsel araştırmalar, araştırmacıların bu tahminlerinin objektif bir şekilde test edilmesi prensibi üzerine inşa edilir. Sosyal bilimlerde genelde maddî hipotezlerden yola çıkılarak, anlamlılık testi ilkelerine uygun istatistiksel hipotezler geliştirilir ve sınanır. İstatistiksel hipotezlerin ilgili testler ile sınanmasına Farksızlık Hipotezi İstatistiksel Anlamlılık Testi (FHİAT)[83] denilmektedir. İstatistiksel hipotezlerin sonuçlarına göre maddî hipotezlere ilişkin bir karara varılması çokça eleştirilmesine

---

[80] Bu çalışmada analiz birimi bireyler olduğundan, tekrarlı ölçümler analizi kullanıldığından ve veri üretiminde aynı tohumlar kullanıldığından aslında özelliklerin dağılımının sadece şekli değil kendisi de sabit tutulmuştur. Bir başka değişle tüm deney koşulları aynı 300 sanal kişi üzerinde uygulanmıştır.

[81] Dikkat edilirse, ayırt edicilik parametresinin (ve madde konumu parametresinin) kendisi değil, dağılımının şekli sabit tutulmuştur.

[82] Bir çalışmada bir faktör belirli bir düzey için sabitlenirken başka bir çalışmada aynı unsur farklı bir düzey için sabitlenebilir. Dolayısıyla ayrı ayrı ele alındığında bu iki çalışma bize söz konusu faktörün bağımlı değişken üzerindeki etkisi hakkında bir bilgi sağlamayacaktır. Ancak aynı konuda yapılan araştırmaları inceleyen bir meta-analiz çalışmasında söz konusu faktör bağlam değişkeni olarak yer alabilir ve bağımlı değişkenle ilişkisi hakkında dolaylı bir bilgiye sahip olunabilir (bağlam değişkeni kavramı için bkz. Fitz-Gibbon, (1984), genel olarak meta-analiz ile ilgili olarak bkz. Durlak, (1995).

[83] Bu kavramın İngilizcesi *Null Hypothesis Statistical Significance Testing*'dir ve *NHSST* şeklinde kısaltılır.

karşın[84] sosyal bilimlerde en sık başvurulan yöntemlerden birisidir. Ancak bu çalışmada FHİAT-merkezli olmayan bir yaklaşım benimsenmiştir.

Bu çalışmanın temel hipotezi; Çok İhtimalli Rasch Modeli ile Derecelendirilmiş Yanıt Modeli arasında tahminleme performansı açısından fark olmadığı iddiasıyla ilgilidir.[85] Bir başka deyişle bağımlı değişkenler birlikte ele alındığında iki ölçme modeli arasında mühim bir fark olmadığı tahmin edilmektedir. Yani bu çalışma temel hipotez açısında ele aldığında Kabul-Destek tipinde bir araştırmadır.

Araştırmanın asıl yapılış sebebini teşkil eden bu öncelikli hipotezin yanı sıra, soru sayısı ve kategori (şık) sayısı ile tahminleme performansı arasında pozitif yönde doğrusal bir ilişki olduğu da iddia edilmektedir. Kategori sayısının artması, denek hakkında daha çok bilgi sağlayacağı için tahminleme performansını da arttıracaktır. Bu zaten bilindiği için bu hipotezin sonucu aslında merak uyandırmamaktadır. Ancak kategori sayısı ile ölçek uzunluğu ve örneklem büyüklüğü arasında nasıl bir etkileşim olduğu bilinmediğinden bu faktörün modele dâhil edilmesi uygun görülmüştür. İfade (soru) sayısının artmasının, bir başka deyişle testin uzatılmasının parametre tahminleme başarısındaki olumlu etkisi olduğu da bilinmektedir. Ancak çok kısa ölçekler için de bu olumlu etkinin geçerli olup olmadığı ve test uzunluğu ile kategori sayısı ve örneklem büyüklüğü gibi faktörler arasında ne gibi bir etkileşimin söz konusu

---

[84] FHİAT ile ilgili eleştiriler ve farklı görüşler ve bu çalışma benimsenen yaklaşım için lütfen Ek-3'deki açıklamaya bakınız.

[85] FYİAT merkezli bir yaklaşım benimsenseydi ve çok değişkenli hipotez yazım formatına uygun bir şekilde hipotez ifade edilmek istenseydi, araştırmanın temel iddiasının farksızlık hipotezi (Stevens, 2009, s. 149'dan esinlenerek);

H_0: "*Aynı anda hem tahminleme yanlılığı, hem de görgül hata varyansı açısından karşılaştırıldığında, Çok İhtimalli Rasch Modeli çerçevesinde yapılan örtük özelliklere ilişkin tahminlemeler ile Derecelendirilmiş Yanıt Modeli çerçevesinde yapılan tahminlemeler arasında istatistiksel olarak anlamlı bir fark yoktur.*"

şeklinde olacaktır. Semboller ile gösterilmesi ise;

$$H_0: \begin{bmatrix} \mu_{11} \\ \mu_{21} \end{bmatrix} = \begin{bmatrix} \mu_{12} \\ \mu_{22} \end{bmatrix}$$

şeklinde olacaktı ki burada μ anakütle ortalamasını, birinci indis değişken numarasını, ikinci indis ise düzey (grup) numarasını temsil edecekti.

olduğu yeterince incelenmemiştir. Çalışmada ayrıca tahminleme performansının, örneklem büyüklüklerine göre farklılık göstereceği de öngörülmektedir.

Bu araştırma dâhilinde temel etkiler ile ilgili olarak spesifik beklentilerin olmasına karşın birçok etkileşim etkisi için bu geçerli değildir ve dolayısıyla araştırmanın bu kısmında keşfedici bir bakış açısı benimsenmiştir. Yalnızca iki etkileşim etkisi ile ilgili belirli bazı beklentiler söz konusudur: (a) Birçok (Monte Carlo) simülasyon çalışmasında madde sayısı ve şık sayısı yerine tahminlenecek madde parametresi sayısı (hatta tahminlenecek parametre sayısının örneklem büyüklüğüne oranı) bağımsız değişken olarak kullanılmıştır (mesela bkz. de Ayala ve Sava-Bolesta, 1999). Araştırmacıların böyle bir tercihte bulunmalarının nedeni tahminlenecek parametre sayısının; test uzunluğu ve şık sayısının bir fonksiyonu olduğu gerçeğidir.[86] Dolayısıyla test uzunluğu ve şık sayısını ayrı değişkenler olarak modele dâhil etmek yerine bileşkesi olan tahminlenecek parametre sayısını değişken olarak kullanmak bağımsız değişken ekonomisi sağlamak amacıyla mantıklı olabilir. Ancak böyle bir tercihte bulunmak, test uzunluğu ve kategori sayısı arasında bir etkileşimin olmadığını varsaymak anlamına gelmektedir. Bu çalışmada bunun aksi düşünüldüğü için, bir başka deyişle test uzunluğu ile kategori sayısı arasında ciddiye alınması gereken düzeyde bir etkileşim etkisi olduğu düşünüldüğü için, bahsedilen faktörler birleştirilmeden modele dâhil edilmiştir. (b) Bunun dışında örneklem büyüklüğünün tahminleme performansı üzerindeki etkisinin iki model için ayrı düzeylerde olacağı öngörülmektedir. Böyle bir beklentinin oluşmasının asıl nedeni modellerin farklı özelliklerinin olması değildir. Modellerin analizlerinin yapıldığı yazılımların kullandığı tahminleme yöntemlerinin özellikleri böyle bir beklenti oluşmasına neden olmaktadır.[87,88]

---

[86] Aslında tahminlenecek madde parametresi sayısı test uzunluğu, şık sayısı ve tahminleme modelinin bir fonksiyonudur. Ama bahsi geçen çalışmalarda model sabit tutulduğundan bu üçüncü faktörün bir etkisi olmamaktadır.

[87] Tahminlenecek parametre sayısı da arttığı için MEO tahminleme yönteminde örneklem büyüklüğünün artmasının olumlu etkisi daha kısıtlıdır.

3.6 Araştırmanın Sınırlılıkları[89]

Bu çalışma bir simülasyon deneyi olduğu için, deneysel tasarımlar ve simülasyon çalışmaları için ayrı ayrı söz konusu olan tüm sınırlılıklar ve dezavantajlar bu çalışma için de geçerlidir.[90] Deneysel çalışmaların en temel sınırlılığı ölçümlerin laboratuar ortamında yapıldığı için sonuçların gerçek hayata (doğal ortama) genellenmesinde ihtiyatlı olunması zorunluluğudur. Bu çalışmanın sanal bir deney olması, bu konuda normalden daha da fazla hassas olmayı gerektirmektedir.

Çalışmada sabit etkiler modeli kullanıldığından sonuçlar yalnızca belirlenen bağımlı değişken düzeyleri için geçerlidir. Örneğin ölçek uzunluğu için 6, 9 ve 12 soruluk olmak üzere üç düzey belirlenmiştir. Elde edilen sonuçlar altıdan az sayıda ifadeden oluşan ölçekler ve on ikiden çok ifadeden oluşan ölçekler için genellenemez.[91]

Simülasyon çalışması ve deneysel tasarımın kullanılmasında kaynaklanan sınırlılıkların dışında bu çalışmada başka kısıtlamalar da söz konusudur: (a) Örneğin eksik veri meselesi araştırmanın dışında tutulmuştur. Yani katılımcıların hepsinin tüm ifadelere eksiksiz yanıt vereceği durumlar simüle edilmiştir. (b) Ölçeklerin aşırı düşük ve aşırı yüksek derecede ayırt edici nitelikte ifadelerden oluşmayacağı varsayılmıştır. Buna bağlı olarak maddelerin ayırt edicilik parametrelerinin 0,8 ile 2 arasında[92] ve

---

[88] Burada sadece beklentilerin söz konusu olduğu etkileşim etkilerine ilişkin hipotezlere değinilmesine rağmen tam faktöriyel bir analiz yapılacağından tüm temel etkiler (4 adet) ve etkileşim etkileri (k faktör sayısı olmak üzere, $[2^k - (k+1)]=11$ adet) test edilecektir.

[89] Araştırma yöntem biliminde; sınırlılık, kısıt, varsayım, ön kabul, ön koşul gibi kavramların anlamları ve birbirlerinden farklılıkları konusunda bir fikir birliği ne yazık ki henüz oluşmamıştır. Bu eksiklik Türkçeye has bir durum da değildir. Bu çalışma çerçevesinde, 'sınırlılık' kavramı ile tasarım, yöntem, analiz tekniği gibi konularda yapılan tercihlerin çalışmanın sonuçlarının genellenebilirliği üzerindeki olası daraltıcı etkileri kastedilmiştir.

[90] Simülasyon çalışmalarının özelliklerine ileriki bölümde değinilecektir.

[91] Hatta 9 sorulu ölçekler dışında, 6 ile 12 soru arasında bir uzunluğa sahip ölçekler için dahi (ölçek uzunluğu ile bağımlı değişken arasında trend analizinin sonuçlarına göre doğrusal bir ilişki olduğu sonucu çıkmış olsa bile) bir genelleme yapılamaz.

[92] Burada belirlenen değerler tipik araştırmalar için yapılan tavsiyeler ile uyumludur (mesela bkz. Hambleton ve Swaminathan, 1985, s. 36). İlk bakışta, ayırt edicilik

dikdörtgensel dağılıma uygun şekilde değişmesine müsaade edilmiştir. (c) İfadelerin zorlukları ile kişilerin yetenek düzeylerinin uyumsuz olduğu durumlar araştırmaya dâhil edilmemiştir (mesela çok yetenekli bir gruba çok kolay ifadelerden oluşan bir ölçeğin uygulanması gibi durumlar araştırma dışında tutulmuştur).

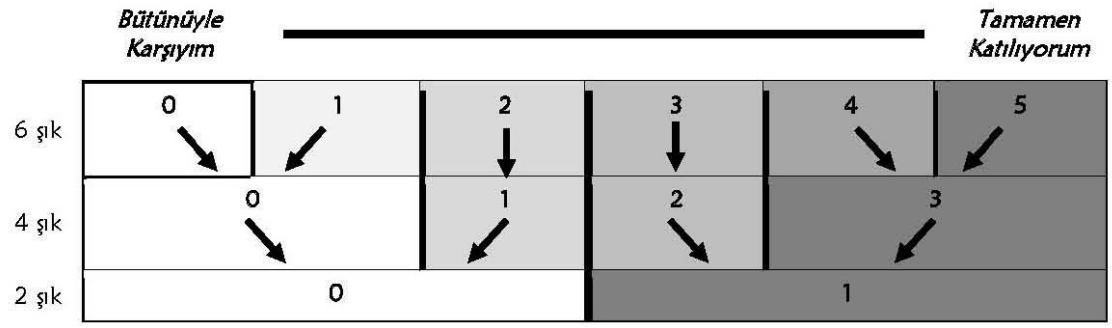
Bu çalışmayla ilgili en ciddi varsayım şudur: Sanal denekler, şık sayısı değişse bile aynı soruya tutarlı ve belirli bir sistem çerçevesinde cevap verecek şekilde bir düzenleme yapılmıştır. Altı şıklı durum baz alınarak farklı şıklar arasında geriye doğru deterministik bir ilişki[93] olduğu kabul edilmiştir. Buna bağlı olarak sadece altı şıklı sorular için yanıtlar üretilmiştir. Dört ve iki şıklı sorular için cevaplayıcının belirli bir mantığa göre yanıt vereceği kabul edildiğinden bu şık sayıları için veri üretim mekanizması çalıştırılmamış, yerine altı şıklı sorular için üretilen yanıtlardan yola çıkılıp yeniden kodlama yapılarak bu sorular için gereken veriler elde edilmiştir. Farklı alternatifler olmasına karşın bu çalışmada aşağıdaki şekildeki (Şekil 3.1.) gibi bir yeniden kodlama uygulanmıştır. Bu şekilde dört ve iki şıklı sorular için yeniden veri üretmek yerine yeniden kodlama yapmaktaki amaç veri üretimindeki tesadüfîliğin etkisini ortadan kaldırmaktır.[94]

---

parametresinin ne kadar yüksek olursa o kadar iyi olacağı kanısına varılabilir. Ancak unutmamak gerekir ki bu parametre değeri arttıkça, ifadenin faydalı olduğu yetenek alanı daralmaktadır. Bu durum, 'bilgi fonksiyonu' konusu ile de ilgilidir. Bilgi fonksiyonu hakkında bilgi için bkz. Hambleton, Swaminathan ve Rogers (1991, 6. Bölüm).

[93] Başka bir değişle, bir cevaplayıcının altı şıklı bir soruya nasıl yanıt verdiği bilinirse, aynı sorunun dört şıklı veya iki şıklı versiyonunun sorulması durumunda o cevaplayıcının ne yanıt vereceğinin şaşmaz bir şekilde bilinebileceği varsayılmıştır.

[94] Ancak bu da insanların kendi tutumlarının yoğunluklarıyla ilgili tam bir tutarlılığa sahip olduklarını kabul etmek anlamına gelmektedir ki bu gerçekçi olmayan bir temennidir Burada, ideal durum test edilmeye çalışılmıştır. İdeal durumda ile gerçek durum arasındaki fark, faktörlerin her düzeyi için aynı olacağından bunun modellerin mukayesesine ilişkin varılacak sonuçlar üzerinde önemli bir fark yaratmayacağı düşünülmektedir.



Şekil 3.1. Yeniden Kodlama Çizelgesi

Genel olarak, istatistiksel açıdan çok daha güçlü olduğu için standart analiz yerine mümkün olan her koşulda tekrarlı ölçümler tasarımı (ve analizleri) kullanılması tavsiye edilmektedir. Ancak bu tasarım türünün de bazı sakıncaları bulunmaktadır (Keren, 1993). Bu çalışma aynı zamanda tekrarlı ölçümler (birimler-içi) tasarımına sahip olmasına karşın tipik bir tekrarlı ölçümler tasarımının kısıt ve dezavantajlarını barındırmamaktadır. Bunun nedeni tekrarlı ölçümler tasarımının sakıncalarının tamamıyla beşeri unsurlardan kaynaklanmasıdır.[95] Bu çalışmada gerçek veriler yerine sunî veriler kullanıldığı için söz konusu sakıncalar bu araştırma için geçerli değildir.[96]

---

[95] Birimler-içi tasarıma ilişkin potansiyel tehlikelerin başında 'devam eden tesir' [*carry over effect*] sorunu gelmektedir. Ayrıntılar için bkz. Myer ve Wells (2003, s. 345), Maxwell ve Delaney (2004 s. 561) ve Kline (2009, s. 82). Ayrıca, Birimler-içi, Birimler-arası tasarımların mukayesesi için bkz. Keren (1993).

[96] Suni verilerin kullanılması sadece tasarıma ilişkin sorunları ortadan kaldırmaktadır. Tekrarlı ölçümlerin analizinde kullanılacak istatistiksel yöntemlere ilişkin varsayımlar ve kısıtlamalar bu çalışma için de geçerlidir.

3.7 Simülasyon Yöntemi[97] ve Veri Üretimi

Simülasyon yöntemi, psikometrik araştırmalarda en sık kullanılan metotlardan birisidir. Gerçek parametre değerleri, suni verilerin üretildiği çalışmalar dışında asla bilinemeyeceği için model karşılaştırmaları, varsayım ihlallerinin sonuçlar üzerindeki etkileri (analizlerin berklik düzeylerinin tespit edilmesi) gibi konularla ilgili araştırmalar için simülasyon çalışması yapmak adeta bir zorunluluktur.[98] Simülasyon çalışmalarının aslında birer deneysel çalışma olduğu ve deneysel çalışmaların sonuçlarının geçerli kabul edilebilmesi için sağlanması gereken şartların (bkz. Kline, 2009, 4. bölüm) hepsinin simülasyon çalışmaları içinde geçerli olduğu unutulmamalıdır. Ayrıca elde edilen bulguların mutlaka istatistiksel analiz yöntemleri ile incelenmesi (Anova, regresyon analizi gibi) gerekmektedir (Harwell, Stone, Hsu ve Kirisci, 1996). Birçok simülasyon çalışmasında ya araştırma istatistiksel analiz yapılmaya engel olacak şekilde tasarlanmıştır (mesela yinelemesiz bir model kurularak) ya da sonuç çıkarıcı analizler yapılabilinecek olmasına karşın sadece betimleyici istatistikler ve grafikler ile yetinilmiştir.[99]

Bu çalışma çerçevesinde; (a) sahip oldukları örtük özellik bakımından standart normal dağılıma sahip bir bireyler anakütlesinden tesadüfî örnekleme yapıldığı, (b) zorluk parametresi açısından standart normal dağılım özelliği gösteren (ve aynı örtük yapıyı ölçen soruların oluşturduğu) bir ifade havuzundan (veya evreninden) tesadüfî

---

[97] Simülasyon yöntemi ile Monte Carlo Çalışması (veya Simülasyonu) kavramları genellikle (ve bilhassa Test Kuramı literatüründe) eş anlamlı olarak kullanılmaktadır. Dolayısıyla bu çalışma da bir Monte Carlo simülasyon çalışması olarak kabul edilebilir. Ancak teknik açıdan doğru olan yaklaşım, Monte Carlo simülasyonu kavramının yalnızca, görgül örnekleme dağılımlarının incelenmesi için suni veri üretilen çalışmalar için kullanılmasıdır. Başka nedenlerle de olsa (mesela anakütle parametrelerinin incelenmek istenmesi gibi nedenlerle) içinde tesadüfi veri üretiminin yapıldığı çalışmalara ise (sadece) simülasyon çalışması denilmesi daha yerinde olacaktır.

[98] Ancak simülasyon çalışmalarına yalnızca problemlere analitik çözümler getirilemediği durumlarda başvurulması gerektiği yönünde ikazlar vardır (Harwell ve diğ., 1996)

[99] Sosyal bilimlerde simülasyon yöntemleri ile ilgili olarak bkz. Bandalos (2006), Mooney (1997) ve en önemlisi (Harwell ve diğ., 1996).

örnekleme yapıldığı ve (c) bu ifadelerin ayırt edicilik parametrelerinin 0,8 ile 2,0 arasında dikdörtgensel (üniform) bir dağılıma sahip olduğu varsayılmış ve bu özelliklerde veriler üretilmiştir. Veri üretimi için WinGen2 (Han ve Hambleton, 2007) ücretsiz yazılımı kullanılmıştır.[100]

Çalışma için, tüm faktörler açısından 'hiyerarşik yuvalı' veri setleri kullanılmıştır. Bu çerçevede örneğin kalibrasyon örneklemi büyüklüğü faktörü için, standart normal dağılan örtülü özellik anakütlesinden 300 yetenek parametresi çekilmiştir. Böylece kalibrasyon örneklem büyüklüğünün ilk düzeyi oluşturulmuştur. Sonra bu 300 parametreden 150'si tesadüfi olarak seçilmiş ve ikinci düzey oluşturulmuştur. Bir faktörün en kapsayıcı düzeyi için verilerin üretilip, diğer düzeyler için bu verilerin içinden tesadüfi olarak seçilmesi yöntemi ifade sayısı (ölçek uzunluğu) ve kategori sayısı faktörleri için de aynen uygulanmıştır. Daha sonra bu veri üretim prosedürü dört kez daha tekrarlanmıştır. Elde edilen bu veri setleri Mulitlog 7 ve Winsteps Programları kullanılarak analiz edilmiş ve özellik parametreleri tahminlenmiştir. Daha sonra bu tahminler kullanılarak bağımlı değişkenlere ait değerler elde edilmiştir. Analizlere geçilmeden önce son olarak örneklem büyüklüğünü eşitlemek için 300 sanal kişiden oluşan veri setlerinden, daha önceden belirlenen 150 kişiye ait olan değerler dışındakiler çıkarılmıştır. Bir başka değişle 72 veri setinin yarısı 150 kişi ile yapılan tahminleme değerlerinden oluşurken diğer yarısı 300 kişi ile yapılan tahminleme değerlerinden 150'sinin seçilmesi ile oluşturulmuştur.

3.8 Araştırmada Kullanılan Veri Analiz Yöntemi

Önsözde de hatırlatıldığı gibi bilimsel araştırmalarda, araştırma tasarımı ile veri analizi (ve ölçüm) birbirlerinden ayrı olarak düşünülemez. Araştırma için tercih edilen tasarım özellikleri, kullanılacak analiz türlerini de belirleyecektir. Bu çalışmada tekrar eden ölçümler içeren (sanal) deneysel bir tasarım tercih edildiği için uygulanması en mantıklı analiz türü tekrarlı ölçümler varyans analizi olmuştur.

Varyans analizi kavramı, tek bir analiz türünün değil bir analiz teknikleri ailesinin genel adı için kullanılmaktadır. Genel olarak varyans analizini 3 açıdan

---

[100] Bu yazılıma http://www.umass.edu/remp/software/wingen/downloadsF.html adresinden erişilebilir.

sınıflandırılabilir. İlk olarak, bağımsız değişken sayısına, yani incelenen faktör sayınına göre (a_1) Tek Yönlü Varyans Analizi ve (a_2) Çok Yönlü (Faktörlü) Varyans Analizi olmak üzere iki türden bahsedilebilir. Bağımlı değişken sayısına göre ise (b_1) Tekdeğişkenli ve (b_2) Çokdeğişkenli[101] Varyans Analizi türleri vardır. Ayrıca deneklerin birden fazla kez ölçülüp ölçülmemesine göre (c_1) Standart (Birimler-arası), (c_2) Tekrarlı Ölçümler (Birimler-içi[102]) ve (c_3) Karma Tasarım[103] Varyans Analizlerinden bahsedilebilir[104].

Bu çalışmada iki ilintili (birbiriyle korele) değişkenin dört bağımsız değişkenden etkilenip etkilenmediği aynı denekler üzerinde tekrarlı ölçümler

---

[101] İngilizcedeki '*multivariate*' kavramına karşılık olarak Türkçede 'çokdeğişkenli' terimi yaygın bir şekilde kullanılmaktadır. Aslında,'çok bağımlı değişkenli' terimi daha doğru bir karşılık olacaktır.

[102] Veri analizi bağlamında, İngilizcedeki '*subject*' kavramına karşılık olarak denek yerine birim kavramı kullanılmıştır. Tüm araştırmalar için jenerik denekler-içi ve denekler-arası kavramlarını kullanmak (analiz biriminin bireyden farklı düzeylerde olabilme ihtimali de göz önüne alındığında) yanlış anlaşılmalara neden olabilir. Belirtmekte fayda var ki, buradaki yanlış anlaşılmalara neden olan şey *subject* kavramının denek olarak çevrilmesi değil, İngilizcede '*between-subjects*' ve '*within-subjects*' kavramlarının jenerik isim olarak kullanılmasıdır.

[103] Karma Tasarım [*mixed design*] kavramı ne yazık ki iki farklı durumu anlatmak için kullanıldığından karışıklığa neden olabilmektedir. Gerçekten de uluslararası literatürde karma tasarım kavramı; (a) içinde hem sabit etkiler, hem de tesadüfî etkiler içeren çok yönlü (faktöriyel) çalışmalar kastedilirken de (Maxwell ve Delaney, 2004, s. 474), (b) hem birimler-içi hem de birimler-arası faktörler içeren tasarımlar kastedilirken de (Gamst, Meyers ve Guarino, 2008, s. 361) kullanılabilmektedir. Dahası, tekrarlı ölçümlere tekdeğişkenli yaklaşımın, çokdeğişkenli durumlar için genelleştirilmiş haline de Çokdeğişkenli Karma Model (ÇKM) denilebilmektedir (mesela bkz. Boik, 1988).

[104] Bu türler için genelde şu İngilizce isim ve kısaltmalar kullanılmaktadır. a1: *Oneway ANOVA* a2: *Multiway/Factorial ANOVA* b1: *(Univariate) ANOVA* b2: MANOVA c1: *(Within-Subjects) ANOVA* c2: *Between-Subjects/Repeated Measures/Correlated Groups ANOVA* c3: *Mixed Design ANOVA*. Bunların dışında bağımsız değişkenlerin düzeylerinin belirleniş tarzına göre sabit etkiler [*fixed effects*] ve tesadüfi etkiler [*random effects*] şeklinde bir ayrımdan da bahsedilebilir.

yapılarak incelendiği için analiz metodu olarak Dört Yönlü Tekrarlı Ölçümler Çokdeğişkenli Varyans Analizi kullanılmıştır[105]. Ayrıca, tüm faktörlerin düzeyleri iradî olarak belirlendiği için bu tasarım sabit etkiler model türüne dâhildir.

Genel Doğrusal Model Analizi açısından bu çalışma aynı zamanda tam faktöriyel bir tasarıma sahiptir. Yani faktörlere ilişkin tüm temel etkilerin[106] ve tüm faktör-faktör etkileşimleri analiz edilmiştir.

Bir çalışmadaki faktör sayısının artması, etkileşim sayısını katlanarak artıracaktır. Böyle bir durumda araştırmacı her bir etkileşim etkisini incelenmeye değer bulmayabilir ve analize dâhil etmeyebilir. Bazı durumlarda ise bir olgu salt etkileşim değişkeni olarak düşünüldüğünden (yani bağımlı değişkene doğrudan bir etkisinin olmadığı öngörüldüğünden) o değişkene ilişkin temel etki incelenmeyebilir. Bazense (eğer araştırmacı aşırı muhafazakâr[107] ise) deneyboyu alfa düzeyinin şişmesinden endişe duyulması sonucunda sadece ilgi duyulan etkiler analize dâhil edilir. Gerekçesi ne olursa olsun tüm temel ve etkileşim etkilerinin incelenmediği bu tür çalışmalara faktöriyel olmayan tasarımlı[108] araştırmalar denilmektedir. Bu çalışmada, literatürde

---

[105] Bazı yazarlar, aynı anda hem birden fazla bağımlı değişkeni olan hem de ikiden fazla düzeyde tekrarlı ölçümlerden oluşan analizleri Duble Çokdeğişkenli Varyans Analizi (DÇVA) [*Double Multivariate ANOVA*] şeklinde isimlendirebilmektedir. Ancak bu kavram, tasarım türünün değil analiz yaklaşım tarzının bir adı olarak kullanılabilir. Çünkü yukarda bahsedilen tarzda bir tasarıma sahip bir çalışmada DÇVA yerine Çokdeğişkenli Karma Model (ÇKM) analizi de tercih edilebilir. İstatistik paket programları genelde fabrika ayarı [*default*] gereği iki analizi birden yaparak raporlarlar. Bu çalışmada da iki tür analize dair sonuçlar verilmesine karşın yorumlar DÇVA sonuçları üzerinden yapılmıştır.

[106] Aslında tam faktöriyel tasarımlar, sadece faktörlere ilişkin değil, eşlik değişkelerine [*covariate*] ilişkin tüm temel etkilerin incelenmesini de içerir. Ancak bu çalışmada eşlik değişken bulunmamaktadır.

[107] Literatürde muhafazakâr sıfatı nedense sadece alfa hatasına karşı hassaslık kastedilirken kullanılan bir nitelemedir. Hâlbuki bir müteakip analiz prosedürü [*post hoc procedure*] veya bu prosedürü kullanan araştırmacı alfa hatasına karşı ne kadar muhafazakâr ise beta hatasına karşı da o kadar liberaldir!

[108] Uluslararası terminolojide bir başka sıkıntı da faktöriyel olmayan [*non-factorial*] tasarım kavramına ilişkin bulunmaktadır. Çünkü faktöriyel tasarım kavramı birden fazla faktör (bağımsız değişkeni) olan çalışmalar için kullanılırken faktöriyel olmayan tasarım kavramı

hâkim olan (bkz. Maxwell ve Delaney, 2004, s. 291) "alfa seviyesi kontrol altında tutulurken deneyboyu değil, etkiboyu alfa[109] düzeyi kritik değere eşitlenir" prensibi benimsendiği için, ilgi duyulan duyulmayan tüm etkilerin analiz edilmesinin, deneyin bütünü için geçerli olan alfa hatasını şişirmesinden endişe edilmemiştir ve tam faktöriyel tasarım kullanılmıştır[110].

İkiden fazla düzeyden oluşan faktörlerin umumî (tümel) testlerinin istatistiksel olarak anlamlı çıkması durumunda müteakip analizlerin yapılması rutin bir uygulamadır. Ancak bu çalışmada hiç birimler-arası değişken olmadığı için (bir başka deyişle bu çalışma bütünüyle birim-içi tasarıma sahip olduğu için) müteakip analiz yapılamamaktadır. Bunun yerine üç düzeyden oluşan test uzunluğu ve kategori sayısı değişkenleri için planlı mukayeseler prosedürü uygulanmıştır. Farklı alternatifler olmasına karşın (mesela Helmert ve fark prosedürleri gibi) bu çalışma için en uygun olduğu düşünüldüğünden polinom planlı mukayese prosedürünün (bazen trend analizi de denilmektedir) uygulanması tercih edilmiştir. Hatırlanacağı üzere araştırmanın temel hipotezlerinden ikisi; (a) test uzunluğu ve (b) kategori sayısı ile parametre tahminleme performansı arasında (ayrı ayrı) pozitif yönde doğrusal bir ilişkinin olduğuna dairdir. İlişkinin olup olmadığı ve yönü umumî test ile belirlenebilirken, eğer ilişki var ise bunun doğrusal mı eğrisel mi olduğu ancak polinom planlı mukayesesi ile test edilebilir. Bunu yapabilmek için ise bağımsız değişkenlerin metrik veri olması ve belirlenen düzeylerin aralıklarının eşit olması gerekmektedir[111] (Green, Salkind ve Akey, 1997, s. 215). Belirlenen düzey sayısı ise kaç

---

sadece bir tane faktörü olan çalışmalar için değil, çok faktörlü ama tüm etkilerin analiz edilmediği tasarımlar kastedilirken kullanılmaktadır.

[109] Deneyboyu hata oranı [*experimentwise*] (α_{EW}), etkiboyu hata oranı [*familywise*] (α_{FW}), mukayesebaşı hata oranı [*per-comparison*] (α_{PC}) kavramları için bkz. Maxwell ve Delaney, (2004, s. 193-6).

[110] Aslında burada bahsedilen hata türleri FHİAT-merkezli bir yaklaşıma sahip kişileri daha çok ilgilendirmektedir. Cohen (1994) gibi FHİAT karşıtı yazarlar, alfa hatasıyla ilgili yapılan düzeltmelere de karşı çıkmaktadırlar.

[111] Bu prosedürün uygulanmasına daha araştırmanın tasarım aşamasında karar verildiğinden test uzunluğu ve şık sayısı değişkeninin düzeyleri arasındaki aralıklar eşit olacak şekilde belirlenmiştir.

farklı sınamanın yapılabileceğini belirler. Buradaki değişkenler üçer düzeyden oluştuğu için sadece ikişer mukayese yapılabilir. Bir başka deyişle yalnızca doğrusal mı yoksa ikinci dereceden eğrisel mi (tek bükümlü) bir trendin daha anlamlı olduğu test edilecektir. Tüm bu analizleri yapabilmek için SPSS programının Genel Doğrusal Model (GLM) prosedürü kullanılacaktır.[112]

3.9 Araştırma Sürecinin Adımları (Araştırmanın Akış Planı)

Bu bölümde tüm araştırma sürecinin kısa bir özeti maddeleştirilerek sunulmuştur. Bağımlı ve bağımsız değişkenlerin, bağımsız değişkenlerin düzeylerinin ve diğer simülasyon koşullarının neler olacağı (daha önce yapılan gerçek görgül araştırmalar da göz önünde bulundurularak) belirlendikten sonra adım adım şu işlemler yapılmıştır.

a. WinGen2 yazılımı kullanılarak belirlenen nitelikte 300 kişi için örtük yetenek parametresi üretilmiştir.
b. Zorluk düzeyleri standart normal dağılan, ayırt edicilik parametreleri 0,8 ile 2 arasında dikdörtgensel dağılan ve altı şıktan oluşan 12 ifade, (Genelleştirilmiş Kısmi Puanlar Modeline göre) üretilmiştir.
c. Üretilen parametrelerinin belirlenen dağılım özelliklerine uygunluğu (dağılım grafiği incelenerek) görsel olarak kontrol edilmiştir.
d. Elde edilen sanal kişilerin sanal ifadelere vereceği cevaplar simüle edilmiştir.
e. Yanıtlar bu haliyle kaydedilmiş ve tek boyutluluk varsayımını ihlal edip etmediği kontrol edilmiştir.
f. Elde edilen yanıtlar Winsteps (Çok İhtimalli Rasch Modeline göre) ve Multilog (Derecelendirilmiş Yanıt Modeline göre) programlarıyla analiz edilmiş, sanal kişilerin örtük özellikleri tahminlenmiş ve standardize edilmiştir.
g. d basamağında elde edilen yanıtlar, yeniden kodlanarak 4 şıklı ifadelere verilen yanıtlar oluşturulmuştur ve f basamağı tekrar edilmiştir.

---

[112] Bu tarz bir tasarıma sahip olan araştırmalarda alternatif olarak SPSS'in Manova Prosedürü de kullanılabilmektedir. İki prosedürün mukayesesi için bkz. SPSS Statistics 17.0 Command Syntax Reference (2007, s. 817-8).

h. d basamağında elde edilen yanıtlar yeniden kodlanarak 2 şıklı ifadelere verilen yanıtlar oluşturulmuştur ve f basamağı tekrar edilmiştir (Ancak Multilog programında tahminlemede kullanılacak model iki Parametreli İki İhtimalli Modele çevirilmiştir. Winsteps için böyle bir düzeltmeye gerek olmamaktadır).
i. d basamağında elde edilen yanıtlar, tesadüfî olarak seçilen 3 soru çıkarılarak tekrar ilgili programlarla ayrı ayrı tekrar analiz edilmiştir.
j. Soru sayısı azaltılmış bu veri setine g ve h basamakları tekrar uygulanmıştır.
k. e basamağında elde edilen yanıtlar, i basamağında çıkarılan 3 soruya ilaveten tesadüfî olarak seçilen 3 soru daha çıkarılarak ilgili programlarla ayrı ayrı tekrar analiz edilmiştir.
l. Böyle 300 kişinin özellik düzeylerine yönelik, dokuz farklı durum için iki ayrı tahminleme değeri elde edilmiştir.
m. Eldeki 300 sanal kişinin sadece (tesadüfî olarak seçilen) 150'sine ait olan iki ayrı tahminlemeleri kaydedilmiştir.
n. d ile m arasındaki tüm basamaklar tekrar edilmiş ancak bu sefer sadece m basamağında belirlenen 150 sanal kişi analize tâbi tutulmuştur. Her farklı durum için bu 150 kişiye ait iki ayrı tahminleme değeri elde edilmiştir. Dolayısıyla 150 kişinin 18 farklı durumda yapılan ölçümlere ilişkin iki farklı tahminlemesi elde edilmiştir.
o. b basamağından o basamağına kadar olan tüm adımlar 4 kez daha tekrar edilmiştir.
p. Eldeki gerçek özellik değerleri ve ÇİRM ve DYM çerçevesinde her bir koşul için ayrı ayrı toplam 5 kez tahminlenen özellik değerleri kullanılarak her bir sanal kişinin her farklı seans (koşul) için Tahminleme Yanlılığı ve Görgül Hata Varyansı hesaplanmıştır.
q. Dolayısıyla 150 x 72'lik bir veri matrisi elde edilmiştir. (2 bağımlı değişken x 2 model x 2 örneklem büyüklüğü x 3 test uzunluğu x 3 şık sayısı=72)
r. Veri matrisi SPSS programına aktarılmış, önce analize ilişkin varsayımlar kontrol edilmiş, gerekli düzenlemeler yapılmış (dışadüşen değerler analizden çıkarılmış, normalleştirme dönüşümü uygulanmış) ve daha sonra da Genel Doğrusal Model-Tekrarlı Ölçümler prosedürü uygulanarak veriler analiz edilmiştir.

4. VERİ ANALİZİ VE BULGULAR

Bu bölümde sırasıyla önce MYK'na ilişkin varsayımlar incelenecek, ardından tekrarlı ölçümler varyans analizinin varsayımları incelenecektir. Daha sonra da verilerin analizine geçilecektir.

4.1 Varsayımların Kontrol Edilmesi

İstatistiksel Analizlere geçilmeden önce, yorumların geçerli olarak kabul edilebilmesi için ilgili varsayımların sağlanıp sağlanmadığının kontrol edilmesi gerekmektedir. Bu araştırma çerçevesinde sonuçları etkileme olasılığı bulunan iki tip varsayımdan bahsedilebilir. Bunlar; Madde Yanıt Kuramı Modellerine İlişkin Varsayımlar ve Veri Analiz Yöntemine İlişkine Varsayımlardır.[113]

4.1.1 Madde Yanıt Kuramı Modellerine İlişkin Varsayımın Kontrolü

MYK ile ilgili varsayımların tanıtıldığı bölümde bahsedildiği üzere modellerle ilgili iki temel varsayım bulunmaktadır. Bunlardan ikincisi olan koşullu bağımsızlık varsayımı, ölçek tasarımı ve uygulanışı ile alakalı olduğu için (ve bu bir simülasyon çalışması olduğundan tek başına bu varsayımın ihlalinin mümkün olmadığı için) burada yalnızca tekboyutluluk varsayımına ilişkin iki analize yer verilmiştir.[114]

---

[113] Aslında, bir çalışma için sunî verilerin üretilmesi durumunda, üretilen verilerin belirlenen özellikleri taşıyıp taşımadığının da test edilmesi tavsiye edilmektedir (Harwell, ve diğ., 1996). Bu bir bakıma kullanılan bilgisayar yazılımının istenildiği gibi çalışıp çalışmadığını test etmek şeklinde de yorumlanabilir. Bu tavsiye, özellikle yazılımların bireysel programcılar tarafından geliştirildiği zamanlar için yerinde olabilir ancak günümüzde programların, (piyasaya sürülmeden önce farklı kişiler ve kurumlar tarafından defalarca test edildiği gerçeğinden yola çıkarak) belirlenen ölçütlere göre veri üretip üretmediğinin sınanmasına gerek olmadığı düşülmüştür. Yine de bu çalışmada, verilerin belirlenen dağılımlara uygunluğu (herhangi bir test yapılmamasına karşın) görsel olarak incelenmiştir.

[114] Esasında simülasyon çalışmalarında tekboyutluluk varsayımının ihlali de oldukça düşük bir olasılıktır. Buna rağmen, veri üretiminde kullanılan WinGen2 yazılımının beklenildiği gibi yanıtlar üretip üretmediğini kontrol etmek amacıyla bazı tekboyutluluk analiz yöntemleri uygulanmıştır.

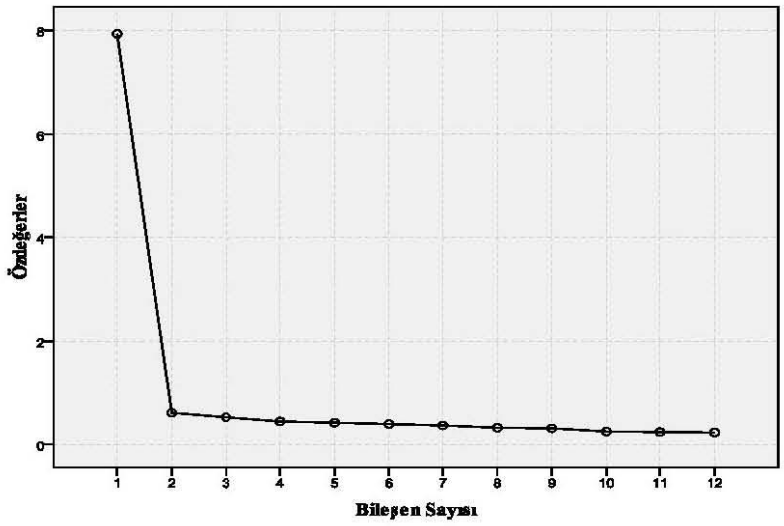
Tekboyutluluk varsayımının kontrol edilmesinin bir yolu ardışık özdeğerler grafiğinin[115] incelenmesidir (Orlando, erişim tarihi, 15 Eylül 2009). Bu yönteme göre grafikte "dirsek" görüntüsünün oluştuğu (yani özdeğerlerde ani düşüşün olduğu nokta) bileşen sayısının bir eksiği verilerin kaç farklı bileşenden oluştuğunu ifade eder. Bu yöntemle verilerin incelenmesiyle elde edilen sonuçlar, verilerin tekboyutluluğunu destekler niteliktedir.[116] Aşağıdaki grafikte (Şekil 4.1.), bu çalışma dâhilindeki en kapsayıcı düzey kombinasyonuna (Kalibrasyon Grubu Büyüklüğü=300, İfade Sayısı=12 ve Kategori Sayısı=6) ait verilere uygulanan ardışık özdeğerler analizinin sonuçları gösterilmiştir.[117]

---

[115] İngilizcesi [*scree plot*] olan bu grafiğe, 'yamaç-birikinti grafiği' (Şencan,2005) de denilebilmektedir.

[116] Her ne kadar birçok araştırmacı tarafından bu ve benzeri Klasik Test Kuramı çıkışlı tekboyutluluk analiz yöntemleri (mesela Keşfedici Faktör Analizi) kullanılsa da, bu yöntemlerin ham verileri eşit aralıklı veri olarak kabul etmek gibi ciddi bir kusuru bulunmaktadır. Dolayısıyla MYK (ve de özellikle Rasch Modeli) kullanılırken tekboyutluluğun bu yöntemlerle test edilmesi kendi içinde bir çelişki barındırmaktadır. Bunların yanı sıra önerilen, doğrusal olmayan faktör analizi ile tekboyutluluk varsayımının incelenmesi ise bu analizi çok ihtimalli ifadeler için yapabilen bir ticari yazılımın henüz geliştirilmediği için yapılamamıştır [NOHARM (Fraser ve McDonald, 2003) gibi yazılımlar sadece iki ihtimalli ifadeler için kullanılabilmektedir].

[117] Bu grafik sadece birinci yinelemenin verilerine aittir. Diğer yinelemelerde de çok benzer sonuçlar elde edildiği için grafikleri burada verilmemiştir.



Şekil 4.1. Ardışık Özdeğerler Grafiği

MYK bakış açısına göre uygulanması en mantıklı tekboyutluluk test prosedürlerinden birisi standartlaştırılmış artık değerler (hata değerleri) arasındaki korelasyonların temel bileşenler analizi (TBA- bu yöntemle ilgili ayrıntılar için bkz. Bartholomew, Steele, Moustaki ve Galbraith, 2002) ile incelenmesidir.[118] Bu yönteme göre hata değerlerine uygulanan TBA sonucu elde edilen birinci bileşenin açıkladığı varyansın ikiden küçük olması beklenir. Ayrıca bu bileşenin açıkladığı varyansın yüzdesinin, maddelerin açıkladığı varyansın yüzdesinden az olması gerekmektedir. Aksi durumun olası ikinci bir boyutun varlığına işaret ettiği kabul edilmektedir (Winsteps, 2009, s. 282-4). Bu çerçevede çalışma dâhilindeki en kapsayıcı düzey kombinasyonuna ait verilere TBA uygulanmıştır. Aşağıdaki tabloda (Tablo 4.1.), tüm yenilemeler için hataların birinci bileşenlerine ilişkin özdeğerler verilmiştir. Bahsedilen kriterlerin sağlanması, ölçeklerin tekboyutluluk varsayımını sağladığı şeklinde yorumlanmıştır.

[118] Dikkate edilirse Klasik Test Kuramındakinden farklı olarak ham verilere değil, hatalara temel bileşenler analizi uygulanmaktadır.

Tablo 4.1. TBA Sonucu Elde Edilen Açıklanan Varyans Yüzde ve Değerleri

Yineleme	Maddelerin Açıkladığı Varyans (%)	Birinci Bileşenin Açıkladığı Varyans (%)	Birinci Bileşenin Açıkladığı Varyans
1	35.3%	3.2%	1,4
2	29.8%	3.4%	1,4
3	33.5%	3.1%	1,4
4	32.1%	3.7%	1,4
5	30.2%	4.1%	1,4

Tekboyutluluk varsayımının sınanmasının ardından, ham verilerden yola çıkarak elde edilen bu örtülü özellik tahminleri kullanılarak, tahminleme yanlılığı ve görgül hata varyansı olarak belirlenen iki bağımsız değişkenin değerleri elde edilmiştir. Bir sonraki aşamada bu bağımsız değişkenlere ait verilerin, kullanılacak istatistiksel analiz yönteminin varsayımlarını sağlayıp sağlamadığı incelenecektir.

4.1.2 Veri Analiz Yöntemine İlişkin Varsayımların Kontrolü

Standart varyans analizinin üç varsayımı bulunmaktadır. Bunlar (Stevens, 2007): (a) her grup (düzey) için bağımlı değişkenin normal dağıldığı, (b) grupların anakütlelerinin varyanslarının eşit olduğu ve (c) gözlemlerin bağımsız olduğudur.[119,120] Çokdeğişkenli varyans analizi için ise bu varsayımların çokdeğişkenli versiyonları, (yani (a) çokdeğişkenli normal dağılım, (b) eş kovaryans matrisler ve (c) gözlemlerin bağımsızlığı) geçerlidir (Weinfurt, 1995).

Tekrarlı ölçümlere çokdeğişkenli yaklaşım için ise eş kovaryans matrisler (birimler-arası bir etki yoksa) ve küresellik varsayımları bulunmamaktadır. Bütünüyle denek-içi türünde bir tasarıma sahip bu çalışmada yorumlar, asıl olarak tek değişkenli tekrarlı ölçümlere çokdeğişkenli yaklaşımın çokdeğişkenli versiyonu olan 'duble çokdeğişkenli varyans analizi' sonuçlarına göre yapıldığı için dikkate alınması gereken

---

[119] Aslında varyans analizinin tek bir varsayımı vardır. O da, $\varepsilon_{ij} \sim NID(0, \sigma_\varepsilon^2)$ şeklinde gösterilebilen, hataların; her bir j anakütlesi için bağımsız ve normal dağıldığı varsayımıdır (Ayrıntılar için bkz. Glass ve Hopkins (1995, s. 402-3) ve Gamst, Meyers ve Guarino (2008, 5. Bölüm). Bu varsayım ile yukarıda verilen üç varsayım pratikte aynı anlama gelmektedir.

[120] Tekrarlı ölçümler varyans analizinin ise yukarıdakilere ilaveten küresellik [*sphericity*] varsayımı da bulunmaktadır.

iki varsayım bulunmaktadır (gözlemlerin bağımsızlığı ve çok değişkenli normal dağılım[121]). Varyans analizinin normal dağılım varsayımına karşı oldukça dirençli (berk) olduğu bilinmektedir. Bu durum aşırı çarpık ve basık/sivri veriler için dahi geçerlidir.[122]

Verilerin dağılımlarına ilişkin incelemelere başlanılmadan önce, varyans analizinin dışadüşen noktalardan aşırı etkilendiği bilindiğinden veri seti bu uç değerlerden arındırılmıştır. Bunun için önce standart değeri -4'den küçük ve +4'den büyük olan noktalar veri setinden çıkarılmıştır. Daha sonra, Mahalanobis uzaklıkları (ayrıntılı bilgi için bkz. Sharma, 1996, s. 44-5 ve Stevens, 2009, s. 106-7) incelenerek Z değerlerinin kontrolü ile tespit edilemeyen dışadüşenler belirlenmiş ve veri setinden çıkarılmıştır. Böylece nihaî örneklem büyüklüğü 150'den 125'e düşmüştür.

Uç değerler elendikten sonra her bir veri setinin (72 adet) normal dağılımdan ne kadar saptığı çarpıklık ve basıklık değerleri kontrol edilerek incelenmiştir. Varyans analizinin; aşırı çarpık (çarpıklık katsayısı=2) ve sivri (basıklık katsayısı=6) veriler için bile dirençli olduğu belirtilmesine (Maxwell ve Delaney, 2004, s. 112) rağmen eldeki veri setinin kabul edilebilir düzeylerin de ötesinde normal dağılımdan farklılık gösterdiği tespit edilmiştir. Dolayısıyla analizlere geçilmeden önce verilere normalleştirme prosedürü uygulanması gerekmiştir. Normalleştirme yöntemleri arasında en iyi sonuçların arksinüs dönüşümünde elde edilmesine karşın bu

---

[121] Çoklu normal dağılımın doğrudan test edilmesinin kolay bir yolu bulunmamaktadır. Bundan dolayı genelde, her bir bağımlı değişkenin normal dağılıp dağılmadığını kontrol etmekle yetinilir. Bu çalışmada da böyle yapılmıştır. Ancak gerçekte, bağımlı değişkenlerin ayrı ayrı normal dağılması çokdeğişkenli normal dağılım için gerekli fakat yeterli değildir.

[122] Varyans analizinin dirençli olduğu gerçeği, sadece FYİAT-merkezli yaklaşımı tercih edenler için rahatlatıcı bir bilgidir. Normallik varsayımın ihlalinin bazı etki büyüklükleri üzerinde (mesela Glass'ın indeksi) ciddi olumsuz etkilerinin olmadığı bilinmesine (Shohov, 2003, s. 20) karşın bu çalışmada kullanılan kısmî eta karelerin ihlallerden nasıl etkilendiği hususunda yapılan bir çalışmaya literatürde rastlanmamıştır. Buna rağmen istatistiklerin, daha çok veri setlerindeki birbirine zıt türde normallikten sapmalarından etkilendiği bilindiğinden ve bu çalışmadaki tüm veri setleri aynı yönde (sağa) çarpık olduğundan sonuçların bu varsayımın ihlalinden ciddi bir şekilde etkilenmeyeceği umulmaktadır. Buna rağmen bu durumun çalışmanın meta-analitik faydasını azaltacağı da bir gerçektir.

normalleştirme yönteminin sonucunda elde edilecek verilerin yorumlanmasının oldukça güç olduğu düşüldüğünden bu metot tercih edilmemiştir. Öte yandan, karekök dönüşümünün uygulanmasının, sonuçların yorumlanmasına hemen hemen hiçbir olumsuz etkisi olmayacağı için (bu çalışmanın bağımlı değişkenlerinin böyle bir dönüştürmeye çok müsait olmasından dolayı), bu normalleştirme yönteminin uygulanmasına karar verilmiştir.[123]

Gözlemlerin bağımsızlığı varsayımı ise istatistiksel olmaktan ziyade tasarımla ilgili bir husustur. Dolayısıyla öncelikle tasarımla ilgili bazı noktalara dikkat edilerek bu varsayımın ihlal edilmesi engellenmeye çalışılır. Bazı nedenlerden dolayı araştırmada bu varsayımın ihlal edilmesi engellenemiyorsa, istatistiksel yöntemler değiştirilerek (mesela hiyerarşik doğrusal modelleme yöntemi kullanılarak) söz konusu ihlallerin olumsuz etkilerinden kaçınılabilir. Bu araştırma, bir simülasyon çalışması olmasından dolayı, gözlemlerin bağımsızlığı varsayımının ihlal edilmesi (aksi istenmedikçe) mümkün değildir.

4.2 Analiz Sonuçları

Tek (bağımlı) değişkenli tekrarlı ölçümler deneylerinde, en az bir bağımlı değişken üç veya daha fazla düzeyden oluşuyorsa araştırmacı iki farklı yaklaşımdan birisini tercih ederek varyans analizi uygulayabilir. Bir başka deyişle tipik tekrarlı ölçümler varyans analizi prosedürünü uygulamak yerine 'tekrarlı ölçümlere çokdeğişkenli yaklaşım' da tercih edilebilir. Bu yaklaşım, ardışık tekrarlı ölçümlerin farkının alınıp bunların bağımlı değişken gibi kabul edilmesi prensibine dayanmaktadır. Böylece aslında tek bir bağımlı değişken ve k düzeyli bir bağımsız değişkenli bir model, k-1 bağımlı değişkenli bir modele dönüştürülmüş olur. Bu yüzden gerçekte model dâhilinde tek bir bağımlı değişken olmasına karşın bu yaklaşıma '(tekrarlı ölçümlere) çokdeğişkenli yaklaşım' denilmektedir. Böyle bir yöntem izlemede temel amaç standart tekrarlı

---

[123] Ancak karekök dönüşümü veri setlerindeki çarpıklığı sadece belli bir düzeyde azaltabilmiştir. Yapılan bir araştırmanın sonuçlarının, meta-analiz çalışmalarında kullanılma ihtimaline karşı araştırmacılardan çalışmalarına ilişkin varsayımların ne derecede ihlal edildiğinin raporlanması beklenilir. Bu çerçevede çarpıklık değerleri Ek 3 Tablo 4 'de verilmiştir.

ölçümler varyans analizinin sağlanması oldukça zor olan küresellik varsayımını atlatmaktır. İki yaklaşımdan[124] hangisinin tercih edilmesi gerektiği ile ilgili tavsiyeler muhteliftir[125] (Stevens, 2009, s. 427).

Bununla beraber, bu çalışmada olduğu gibi modelin orijinalinde birden çok bağımlı değişken varsa durum biraz karmaşıklaşır. Ancak temelde uygulanan prosedür, yukarıdaki yöntemin her bir bağımlı değişken için tekrar edilmesinden ibarettir. Böyle birden çok bağımlı değişkeni olan tekrarlı ölçümler tasarımlarında sözü edilen yaklaşımın tercih edilmesiyle kullanılan analize duble çokdeğişkenli varyans analizi denilmektedir.[126] Bu çalışmada, iki bağımlı değişken olduğu ve tekrarlı ölçümler yapıldığı için bu analiz uygulanmıştır. İstatistiksel anlamlılık testlerini ve etki büyüklüklerini içeren tablo aşağıda sunulmuştur.

Çok yönlü varyans analizi sonuçlarını yorumlamaya genelde temel etkilerden başlanır[127]. Ancak Kategori Sayısı değişkeninin kendisinin yanı sıra parçası olduğu ikinci dereceden etkileşim etkilerinin de etki büyüklüklerinin yüksek çıkması sonuçları yorumlamaya önce bu değişkenden başlanılmasına neden olmuştur (bkz. Tablo 4.2.). Diğer bağımsız değişkenlerle etkileşiminin anlamlı çıkmasının olası nedenleri yakından incelendiğinde sonuçların karmaşıklaşmasının iki şıklı soru tipi düzeyinden kaynaklandığı görülmüştür.

---

[124] Bu iki farklı yöntemin, sadece ikiden fazla düzey içeren faktörlerin temel etkileri için ve bu faktörlerin dâhil olduğu etkileşim etkileri için farklı sonuçlar vereceğini hatırlatılır.

[125] Bu çalışmada iki analizin de yapılıp, rapor edilmesi yönündeki öneriler benimsenmiştir.

[126] Literatürde, standart tekrarlı ölçümler yaklaşımının çokdeğişkenli versiyonuna verilen yerleşmiş bir isim bulunmamaktadır (ancak bazı yazarlar bu durum için Çokdeğişkenli Karma Model terimi kullanmaktadır (mesela Boik,1988). Yukarıdaki isimlendirme tarzına paralel olarak bu yaklaşıma da 'tekli çok değişkenli varyans analizi' denilmesi beklenebilir (ki Page, Braver ve MacKinnon, (2003) gibi bazı yazarlar böyle yapmaktadır) ama bu isim genelde, duble çok bağımlı değişkenli varyans analizinde sonuçların anlamlı çıkması halinde bağımsız değişkenlerin, her bir bağımlı değişken üzerindeki etkilerinin ayrı ayrı incelendiği müteakip analiz prosedürü için kullanılmaktadır.

[127] Bu araştırma, 'bütünüyle birimler-içi' tasarıma sahip olduğu için incelenecek birimler-arası temel etki bulunmamaktadır.

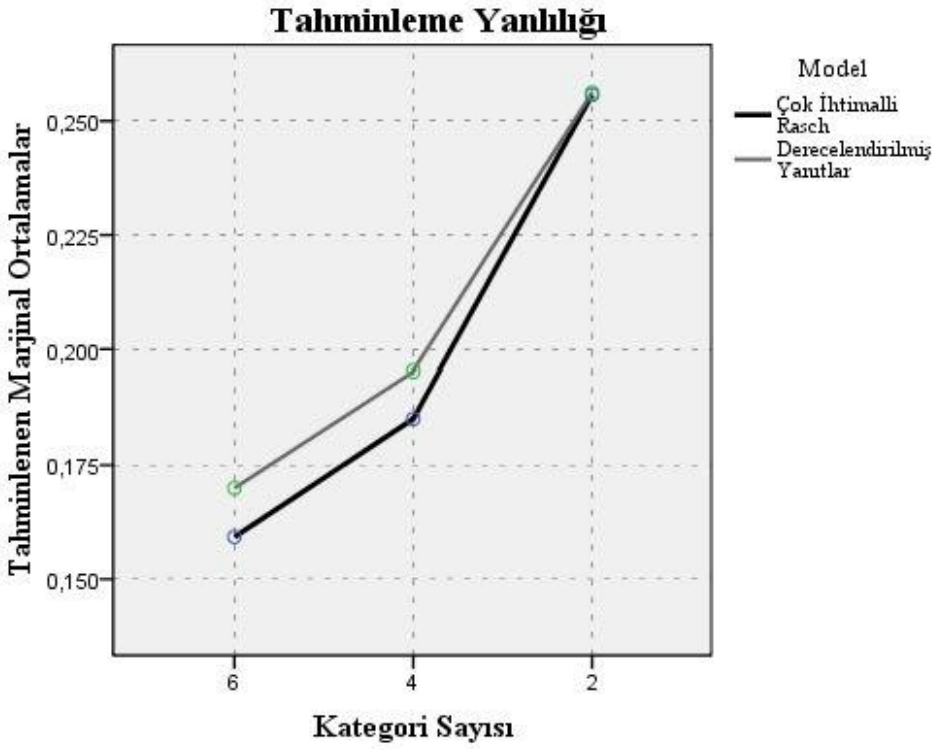
Tablo 4.2. Çok Değişkenli Testler (Tüm Düzeyler İçin)

Etkiler	λ	*F*	*sd*	*p*	η_p^2
Model (M)	,976	1,520	2	,223	,024
Kalibrasyon Örneklemi Büyüklüğü (N)	,863	9,760**	2	<,001	,137
Ölçek Uzunluğu (L)	,325	62,916**	4	<,001	,675
Kategori Sayısı (C)	,264	84,147**	4	<,001	,736
M x N	,928	4,760*	2	,010	,072
M x L	,978	0,675	4	,610	,022
M x C	,944	1,785	4	,136	,056
N x L	,919	2,663*	4	,036	,081
N x C	,920	2,613*	4	,039	,080
L x C	,745	5,009**	8	<,001	,255
M x N x L	,948	1,675	4	,160	,052
M x N x C	,956	1,398	4	,239	,044
M x L x C	,874	2,114*	8	,040	,126
N x L x C	,872	2,155*	8	,036	,128
M x N x L x C	,922	1,233	8	,286	,078

Not. λ: Wilks' Lambda, *F*: Wilks' Lambda'ya ait *F* Değeri, *sd*: Serbestlik Derecesi, *p*: İstatistiksel Anlamlılık Seviyesi, η_p^2: Kısmi Eta Kare Etki Büyüklüğüdür.
* sembolü ,05 düzeyinde, ** sembolü ise ,01 düzeyinde anlamlı *F* değerlerini göstermektedir

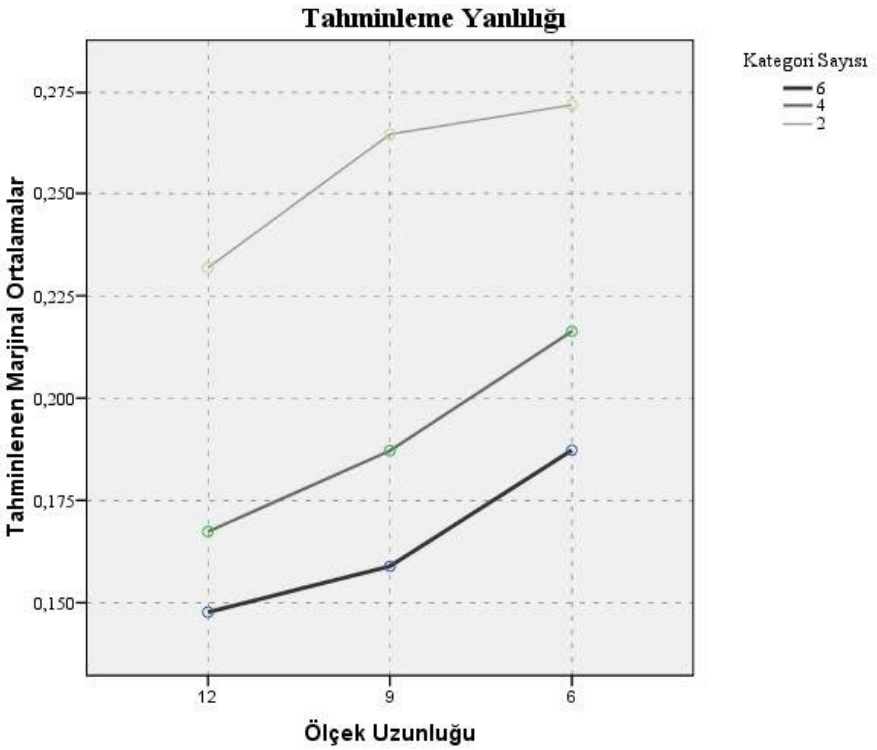
Şekil 4.1.'de tahminleme yanlılığı bağımlı değişkeni üzerinde şık sayısının etkisi, her model için ayrı çizgilerle olmak üzere görsel olarak sunulmuştur. İki model için de kategori (şık) sayısı altıdan dörde düştüğünde yanlılık sadece bir miktar artmaktayken,[128] kategori sayısı dörtten ikiye düştüğünde yanlılık düzeyi sıçrayarak artmaktadır. Şık sayıları arasındaki bu dramatik farkın yanı sıra, görüldüğü gibi iki model arasındaki şık sayısı-yanlılık ilişkisiyle ilgili paralellik, iki şıklı sorularda bozulmaktadır. Bu bozulma, araştırmanın 'modeller arası fark olmadığına' ilişkin temel hipotezini güçlendirir nitelikte olmasına karşın modelin geneline dair yorumları biraz karmaşıklaştırmaktadır.

[128] Bu çalışmadaki iki bağımlı değişken de 'ters kriter' niteliğindedir. Yani bu değişkenler için düşük değerler daha makbuldür.



Şekil 4.2. Tahminleme Yanlılığına İlişkin Model x Kategori Sayısı Profil Grafiği

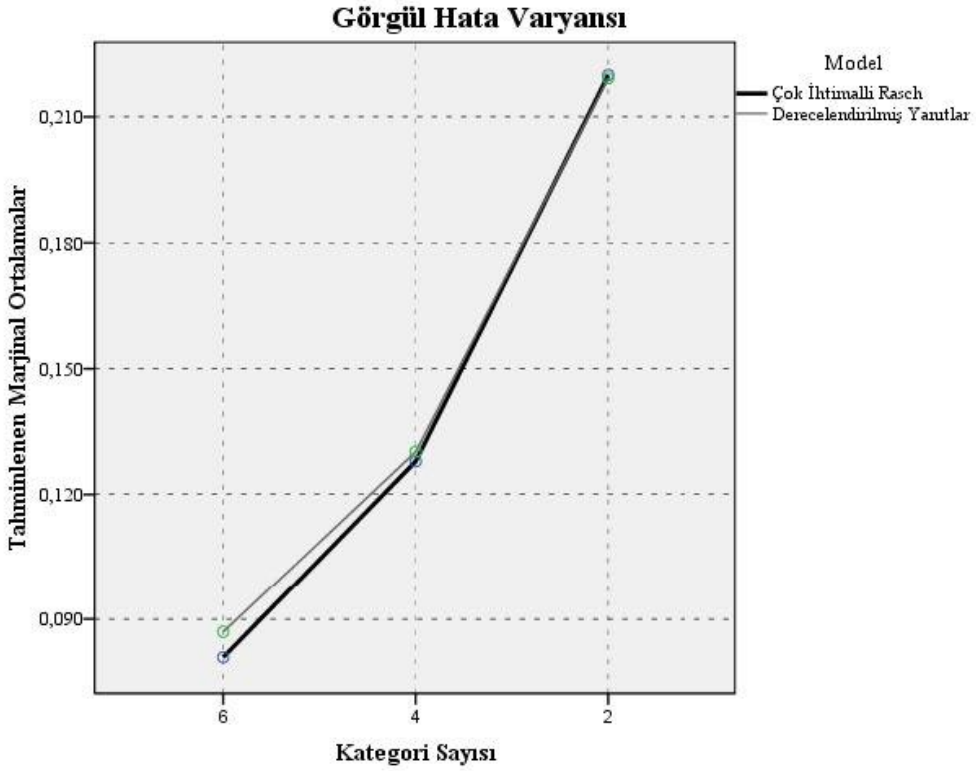
İki ihtimallilik düzeyinin sonuçlar üzerindeki karmaşıklaştırıcı etkisi Şekil 4.2.'de daha bariz bir şekilde görülmektedir. Bu grafikte modellerden bağımsız olarak (yani iki modele ait sonuçların birlikte ele alınması halinde) test uzunluğunun tahminleme yanlılığı üzerindeki etkisi görselleştirilmiştir. Grafikten anlaşılacağı üzere ifade sayısı azaldıkça (ölçek uzunluğu kısaldıkça) tahminleme kalitesi de düşmektedir ve bu durum şık sayısı kaç olursa olsun değişmemektedir. Ancak ifade sayısının artmasının tahminleme yanlılığı üzerindeki iyileştirici etkisi farklı şık sayıları için farklı şekillerde olmaktadır. Mesela dört ve altı şıklı ifadeler için; on iki ifade veya dokuz ifade olması arasında ciddi bir fark yoktur ama dokuz ifade ile altı ifade arasında önemli bir fark vardır. Buna karşın iki şıklı ifadelerde; on iki ve dokuz ifadeli ölçekler arasındaki fark büyükken dokuz ve altı ifadeli ölçekler arasındaki fark cüzidir.



Şekil 4.3. Tahminleme Yanlılığına İlişkin Kategori Sayısı x Ölçek Uzunluğu Profil Grafiği

Ayrıca bu grafikten (Şekil 4.3) ifade sayısı kaç olursa olsun, şık sayısının on ikiden dokuza inmesiyle tahminleme başarısında gerçekleşen düşüşün, şık sayısının dokuzdan altıya indirilmesiyle gerçekleşen düşüşten çok daha fazla olduğu görülmektedir.

İki ihtimalliğin sonuçlar üzerindeki etkisi, şiddeti daha az olmakla birlikte görgül hata varyansı için de geçerlidir. Şekil 4.3.'de kategori sayısının ikiye düşülmesinin görgül hata varyansı düzeyinde nasıl bir sıçramaya neden olduğu ve buna ilaveten modeller arasındaki paralelliğin iki şıklı durumda nasıl bozulduğu görülmektedir. Kategori sayısının ölçek uzunluğu-görgül hata varyansı ilişkisi üzerindeki (etkileşim) etkisi ise Şekil 4.4. grafiğinde gösterilmiştir.



Şekil 4.4. Görgül Hata Varyansına İlişkin Model x Kategori Sayısı Profil Grafiği

Daha önce belirtildiği gibi kategori sayısı değişkeninin düzeyleri arasına iki ihtimalli ifade tipinin de dâhil edilmesindeki amaç bu düzeyin, çok ihtimalli modellerin karşılaştırılmasına temel olmasını sağlamaktı. Ancak iki ihtimalli ifade tipine ilişkin sonuçların diğer düzeylerden bu denli farklı olacağı tahmin edilmemiştir. Bu haliyle, elde edilen bulguların geçerliliğinde herhangi bir sorun olmamasına karşın sonuçların daha kolay yorumlanabilmesini sağlamak amacıyla (sonuçların genellenebilirlik alanını daraltma pahasına) iki ihtimalli soru tipi düzeyi modelden çıkarılarak veriler yeniden analiz edilmiştir. Yeni haliyle, kategori sayısı değişkeni üç yerine iki düzeyde incelendiği için model, 2 (model türü) x 2 (kalibrasyon örneklem büyüklüğü) x 3 (test uzunluğu) x 2 (şık sayısı) tasarımına dönüşmüştür. Yeniden

düzenlenmiş haline ilişkin duble çok değişkenli analiz sonuçları aşağıdaki tabloda (Tablo 4.3.) sunulmuştur[129].

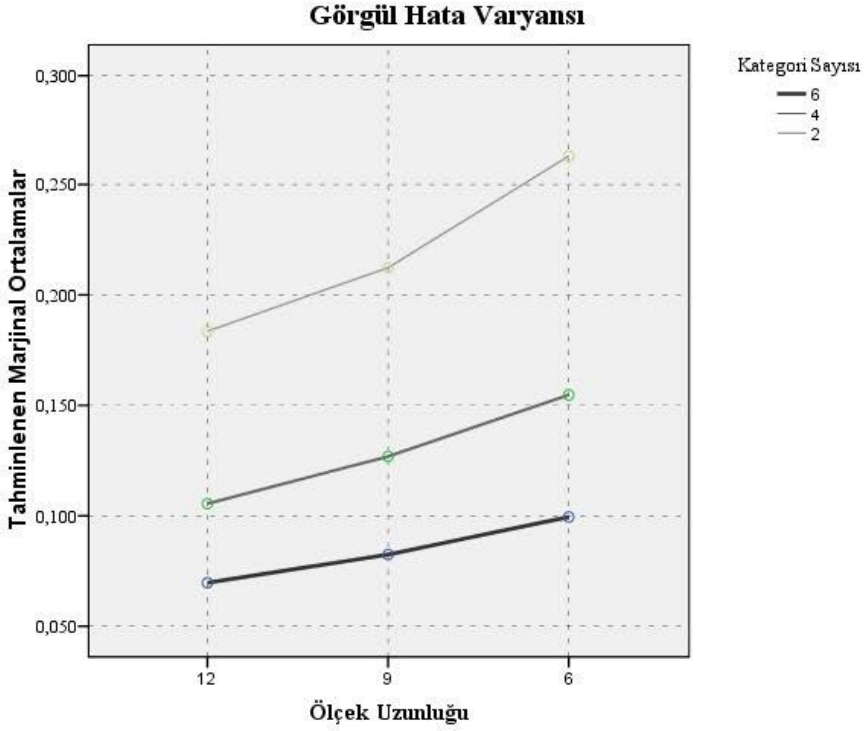
Tablo 4.3. Çok Değişkenli Testler (İki İhtimalli Modeller Hariç)

Etkiler	λ	*F*	*sd*	*p*	η_p^2
Model (M)	,969	1,976	2	,143	,031
Kalibrasyon Örneklemi Büyüklüğü (N)	,805	14,934**	2	<,001	,195
Ölçek Uzunluğu (L)	,399	45,534**	4	<,001	,601
Kategori Sayısı (C)	,511	58,806**	2	<,001	,489
M x N	,900	6,804*	2	,002	,100
M x L	,982	0,553	4	,697	,018
M x C	,992	0,497	2	,609	,008
N x L	,796	7,775**	4	<,001	,204
N x C	,981	1,184	2	,310	,019
L x C	,898	3,436*	4	,011	,102
M x N x L	,924	2,492*	4	,047	,076
M x N x C	,963	2,349	2	,100	,037
M x L x C	,938	2,008	4	,098	,062
N x L x C	,996	0,120	4	,975	,004
M x N x L x C	,954	1,456	4	,220	,046

Not. λ: Wilks' Lambda, *F*: Wilks' Lambda'ya ait *F* Değeri, *sd*: Serbestlik Derecesi, *p*: İstatistiksel Anlamlılık Seviyesi, η_p^2: Kısmi Eta Kare Etki Büyüklüğüdür.

* sembolü, 05 düzeyinde, ** sembolü ise ,01 düzeyinde anlamlı *F* değerlerini göstermektedir.

[129] Analizin yapıldığı SPSS yazılımı standart olarak Wilks'in Lambdası (λ) değerinin yanı sıra Pillai'nın İz, Hotelling'in İz ve Roy'un En Büyük Kökü değerlerini de vermektedir. Ancak bu değerler -farklı olmalarına rağmen değerlere ilişkin istatistiksel anlamlılık düzeyleri ve etki büyüklükleri aynı olduğu için tabloyu sadeleştirmek adına- burada sunulmamıştır.



Şekil 4.5. Görgül Hata Varyansına İlişkin Kategori Sayısı x Ölçek Uzunluğu Profil Grafiği

İstatistiksel etki büyüklükleri (ve istatistiksel anlamlılık testleri incelendiğinde[130]) tüm araştırmanın tüm temel hipotezlerine ilişkin beklentileri destekler nitelikte sonuçlar elde edildiği görülmüştür:

(a) Bu araştırmanın çıkış noktasını olan Çok İhtimalli Rasch Modeli ile Derecelenmiş Yanıtlar Modeli arasında tahminleme kalitesi açısından fark olmadığı yönündeki birinci hipotezle uyumlu sonuçlar elde edilmiştir $(\boldsymbol{\lambda = ,969;\ F(2;123) = 1{,}976;\ p = ,143;\ \eta_p^2 = ,031})$. Etki büyüklüğünün çok düşük olması iki model arasında önemsenecek bir farkın olmadığını göstermektedir.[131] Bu sonuç, çok parametreli modellerin daha iyi sonuçlar verdiğine yönelik hâkim olan görüşün de sarsıldığı

[130] Daha önce de belirtildiği gibi bu çalışmada yorumlar etki büyüklüklerine göre yapılmıştır. Anlamlılık testleri sonuçlarına ilişkin göndermeler yapılmasından bilhassa kaçınılmaya çalışılmıştır. Ancak bu değerlerin yorumlanmasına aşina olan okuyucular için test değerleri tablolarda ve parantez içlerinde verilmiştir.

[131] İlginç olan, önemsenmeyecek düzeyde olan bu farkın ÇİRM lehine olmasıdır.

anlamına gelmektedir. Bir başka deyişle, veri-model uyumunun daha iyi olduğu gerekçesiyle temel ölçme ilkelerinden uzaklaşıp 2 PL modellere yönelmenin, tipik bir örgütsel davranış araştırmacısı için hiçbir anlamı olmadığı sonucuna erişilmiştir.

(b) Büyük kalibrasyon örneklemleriyle yapılan tahminlemelerin küçük örneklemlerle yapılanlara göre daha başarılı olacağına yönelik öngörüyle ilgili hipotezi destekleyen bulgular elde edilmiştir. $(\lambda = ,805; F(2;123) = 14{,}934; p < ,001; \eta_p^2 = ,195)$ Bu faktörün etkisi önemli kabul edilebilecek boyutlarda olmakla birlikte diğer iki temel etkiye göre daha düşük düzeydedir.

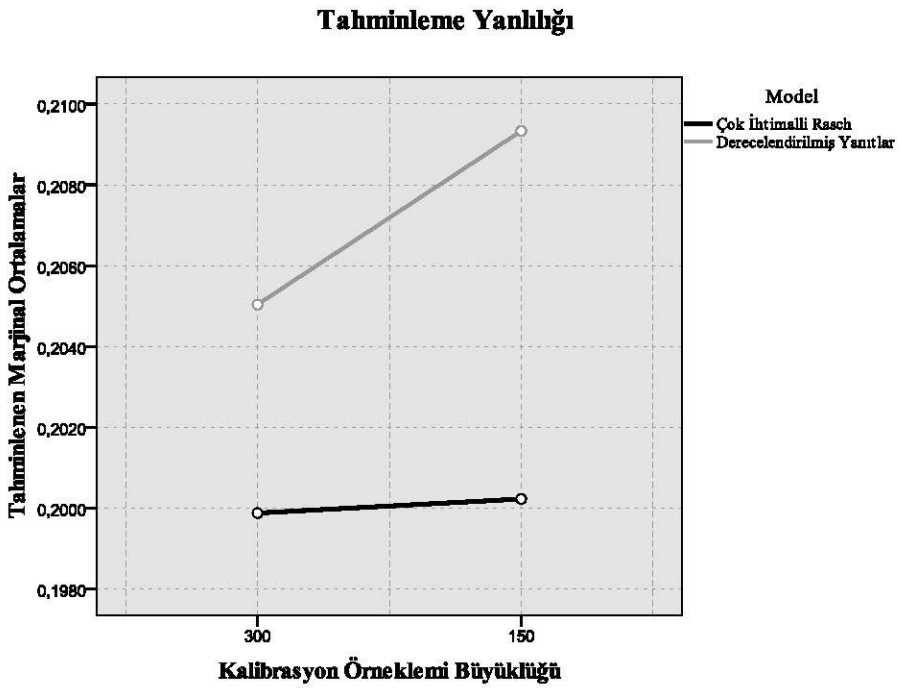
(c) Farklı test uzunluklarına ait tahminleme performansları ile ilgili farksızlık hipotezinin reddedilmesi, istatistiksel farkın[132] olduğunun bir göstergesi olarak kabul edilmiştir. $(\lambda = ,399; F(4;121) = 45{,}534; p < ,001; \eta_p^2 = ,601)$. Tüm etkiler arasında en güçlü olanı, test uzunluğu çıkmıştır. Ayrıca trend analizi sonucunda test uzunluğu-tahminleme yanlılığı ve test uzunluğu-görgül hata varyansı ilişkilerinin doğrusal olduğu sonucu (bkz Tablo 4.6.) bulunmuştur $(p < ,001; \eta_p^2 = ,135 \; ve \; p < ,001; \eta_p^2 = ,572)$

(d) Altı şıkla yapılan ölçümlerle, dört şıkla yapılan ölçümlere göre daha başarılı tahminlemeler yapılacağı şeklindeki hipotez ile de tutarlı olan sonuçlar bulunmuştur $(\lambda = ,511; F(2;123) = 58{,}806; p < ,001; \eta_p^2 = ,489)$.

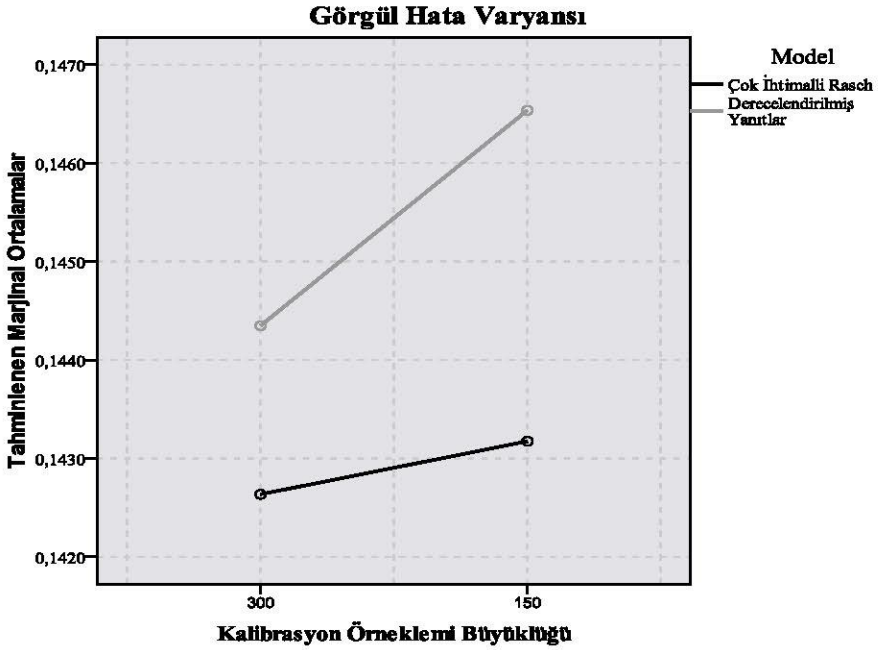
(e) Model türü ile örneklem büyüklüğü arasında olması beklenilen etkileşimsel ilişkiyle ilgili oluşturulan hipotezi kısmen destekleyen sonuçlara ulaşılmıştır $(\lambda = ,900; F(2;123) = 6{,}804; p = ,002; \eta_p^2 = ,100)$. Şekil 4.5 ve 4.6'da bu etkileşim etkisinin ayrı ayrı her iki bağımlı değişken için de geçerli olduğu görülmektedir.

---

[132] Bu çalışmada, Kline'ın (2004, s. 88) önerisine uyularak, 'anlamlılık', hatta 'istatistiksel anlamlılık' kavramlarının kullanılmasından mümkün olduğunca kaçınmaya çalışılmıştır.

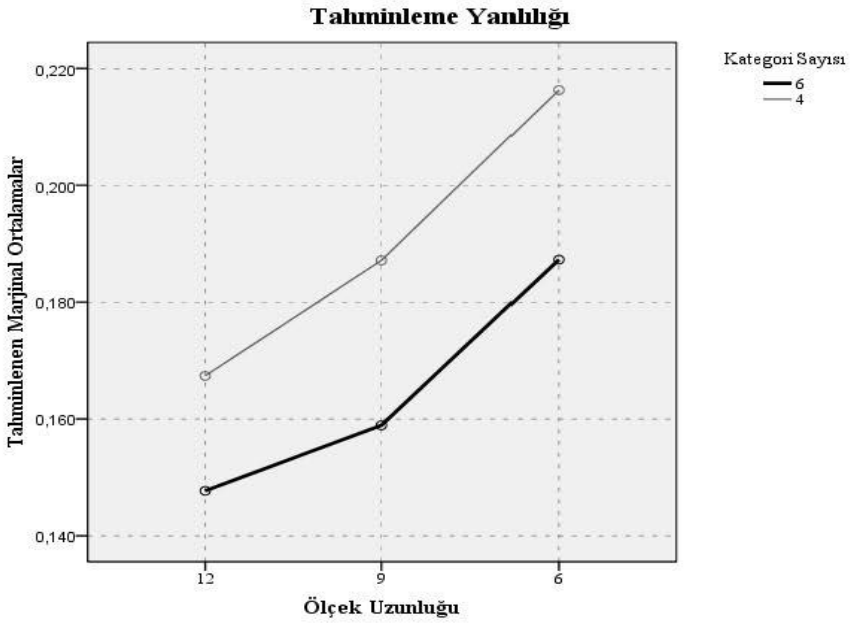


Şekil 4.6. Tahminleme Yanlılığına İlişkin Model x Kalibrasyon Örneklemi Büyüklüğü Profil Grafiği

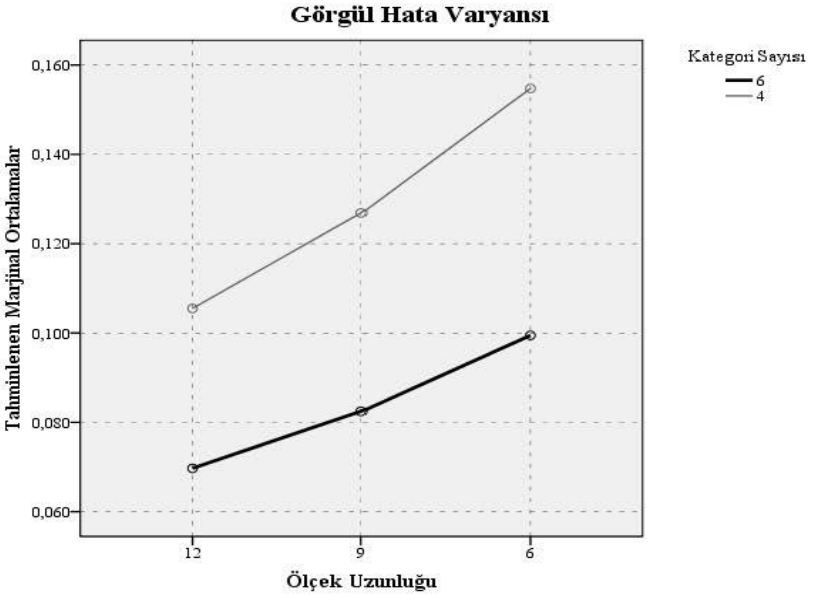


Şekil 4.7. Görgül Hata Varyansına İlişkin Model x Kalibrasyon Örneklemi Büyüklüğü Profil Grafiği

(f) Test uzunluğu ile şık sayısı arasında genel kanının aksine daha karmaşık bir ilişki olduğu yönündeki beklentiler doğrultusunda incelenen Test Uzunluğu x Şık sayısı etkileşiminin (beklenilen kadar yüksek olmasa da) dikkate değer bir etki büyüklüğünün olduğu sonucu bulunmuştur $\left(\lambda = ,102;\ F(2;123) = 3{,}436;\ p = ,011;\ \eta_p^2 = ,102\right)$. Şekil 4-7'deki profil grafiği, Ölçek Uzunluğundaki değişimlerin Tahminleme Yanlılığı üzerindeki etkisinin 6 şıklı ve 4 şıklı sorular için aynı olmadığını göstermektedir. Görgül Hata Varyansı için ise böyle bir etkileşim etkisinin söz konusu olmadığı görülmektedir (bkz. Şekil 4.8.).



Şekil 4.8. Tahminleme Yanlılığına İlişkin Kategori Sayısı x Ölçek Uzunluğu Profil Grafiği



Şekil 4.9. Görgül Hata Varyansına İlişkin Kategori Sayısı x Ölçek Uzunluğu Profil Grafiği

(g) Model x Kalibrasyon Örneklemi Büyüklüğü etkileşimini sadece kısmen destekler nitelikte sonuçlar elde edilmiştir $(\boldsymbol{\lambda = ,900;\ F(2;123) = 7{,}775;\ p = ,002;\ \eta_p^2 = ,100})$. Etki büyüklüğü kabul edilen alt sınırdan yüksek ama beklenilen düzeyden düşük çıkmıştır.

(i) Sonuçlarına yönelik herhangi bir öngörüde bulunulmayan, Ölçek Uzunluğu x Kalibrasyon Örneklemi Büyüklüğü etkileşimine ilişkin yüksek bir etki büyüklüğü elde edilmiştir $(\boldsymbol{\lambda = ,796;\ F(4;121) = 7{,}775;\ p = ,000;\ \eta_p^2 = ,204})$.

(j) Araştırmada hiçbir ikinci veya üçüncü derece etkileşim etkisinin anlamlı çıkmaması beklenmekteydi.[133] Bu beklentilerle tam uyumlu olmayan tek sonuç Model x Kalibrasyon Örneklemi Büyüklüğü x Ölçek Uzunluğu Etkisi ile ilgili sonuçtur. Gerçi bu etkileşim etkisi ile ilgili (*p* değeri "istatistiksel olarak anlamlı" sonuçlar vermesine

[133] Birinci düzey dışındaki etkileşim etkilerinin yüksek çıkmaması genel olarak araştırmacılar tarafından arzulanan bir sonuçtur. Bunun nedeni bu üst derecedeki etkileşim etkilerinin tüm analiz sonuçlarının yorumlanışını zorlaştıracağı gerçeğidir.

karşın) etki büyüklüğü düşük çıkmıştır.[134] Dolayısıyla bu etkinin ne kadar önemli kabul edilmesi gerektiği yoruma açıktır. Bu ikinci düzey etkinin sahi kabul edilmesi halindeki yorumu şu şekilde yapılabilir:

Genel olarak ele alındığında ifade sayısı arttıkça tahminleme kalitesinin arttığı görülmektedir. Ancak bu artışın trendi örneklemin büyük veya küçük olmasına göre değişmektedir. Dahası örneklem büyüklüğünün söz konusu trend üzerindeki etkisi de modelin türüne göre farklılıklar göstermektedir.

Hipotezlere ilişkin yukarıdaki yorumların (ve sunulan istatistiklerin) hepsi duble çok değişkenli varyans analizi yaklaşımı çerçevesinde elde edilmiş bulgulara göre yapılmıştır. Mevcut araştırmanın tasarım özellikleri gereği, tüm hipotezler, daha önce bahsedilen alternatif yaklaşım (çokdeğişkenli karma model) çerçevesinde de değerlendirilebileceği için bu alternatif yaklaşım şekline göre elde edilen bulgular da aşağıdaki tabloda (Tablo 4.4.) sunulmuştur.[135,136]

---

[134] Tekdeğişkenli analizlerde etki büyüklüklerinin yorumlanmasına ilişkin genel tavsiyeler bulunmasına karşın çok değişkenli veya tekrar ölçümlü tasarımlar için hususi değerler belirtilmemiştir. Tekdeğişkenli modellerde η^2 (eta kare) değerinin ,01 olması küçük; ,06 olması orta ve ,14 olması büyük etki büyüklüğünün söz konusu olduğu şeklinde yorumlanmaktadır (Cohen, 1977). Çok değişkenli ve tekrarlı ölçümler tasarımlarının daha güçlü oldukları gerçeğini göz önünde bulundurularak bu çalışmada çokdeğişkenli kısmî $\eta^2 > ,10$ (,06 ile ,14'ün ortalaması) düzeyini ciddiye alınmaya değer etki büyüklükleri için sınır olarak benimsenmiştir.

[135] Bu tabloda tüm etkiler gösterilmemiştir. İki yaklaşım arasında sadece 3 veya üzeri düzeyden oluşan değişkenlere dair temel etkilerin ve bu değişkenlerin parçası olduğu etkileşim etkilerinin değerleri değişeceği için yalnızca bu değerler tablolaştırılmıştır.

[136] Bu sonuçların geçerliliği küresellik varsayımın ne derecede tutturulduğuyla yakından ilgilidir. Bu varsayımın testi için SPSS yazılımı otomatik olarak Mauchly Küresellik Testini yapmaktadır. Ancak bu testin sonuçlarına itibar edilmemesine yönelik literatürde bir fikir birliği oluşmuştur. Yerine düzeltme faktörlerinin uygulanması önerilmektedir. Mauchly Testinin sonuçları ve düzeltme faktörleriyle ilgili istatistikler Ek 3 Tablo 2'de verilmiştir.

Tablo 4.4. Çok Değişkenli Testler (Çok Değişkenli Karma Model)

Birimler-içi Etkiler		Değerler	*F*	*sd*	*p*	η_p^2
Ölçek Uzunluğu (L)	Pillai İzi	,462	37,215**	4	<,001	,231
	Wilks Lambda	,539	44,768**	4	<,001	,266
	Hotelling İzi	,856	52,627**	4	<,001	,300
	Roy En Büyük Kökü	,855	106,013**	2	<,001	,461
Model (M) x L	Pillai İzi	,012	0,761	4	,551	,006
	Wilks Lambda	,988	0,760	4	,551	,006
	Hotelling İzi	,012	0,760	4	,552	,006
	Roy En Büyük Kökü	,012	1,522	2	,220	,012
Kalibrasyon Örneklemi Büyüklüğü (N) x L	Pillai İzi	,147	9,872**	4	<,001	,074
	Wilks Lambda	,853	10,232**	4	<,001	,077
	Hotelling İzi	,172	10,590**	4	<,001	,079
	Roy En Büyük Kökü	,170	21,084**	2	<,001	,145
M x N x L	Pillai İzi	,063	4,058*	4	,003	,032
	Wilks Lambda	,937	4,108*	4	,003	,032
	Hotelling İzi	,068	4,157*	4	,003	,033
	Roy En Büyük Kökü	,067	8,324**	2	<,001	,063
L x Kategori Sayısı (C)	Pillai İzi	,072	4,623**	4	,001	,036
	Wilks Lambda	,928	4,680**	4	,001	,037
	Hotelling İzi	,077	4,737**	4	,001	,037
	Roy En Büyük Kökü	,074	9,201**	2	<,001	,069
M x L x C	Pillai İzi	,038	2,400*	4	,049	,019
	Wilks Lambda	,962	2,409*	4	,048	,019
	Hotelling İzi	,039	2,418*	4	,048	,019
	Roy En Büyük Kökü	,037	4,605*	2	,011	,036
N x L x C	Pillai İzi	,002	0,149	4	,963	,001
	Wilks Lambda	,998	0,148	4	,964	,001
	Hotelling İzi	,002	0,148	4	,964	,001
	Roy En Büyük Kökü	,002	0,290	2	,749	,002
M x N x L x C	Pillai İzi	,022	1,388	4	,237	,011
	Wilks Lambda	,978	1,386	4	,237	,011
	Hotelling İzi	,023	1,385	4	,238	,011
	Roy En Büyük Kökü	,019	2,392	2	,094	,019

Not. *sd*: Serbestlik Derecesi, *p*: İstatistiksel Anlamlılık Seviyesi, η_p^2: Kısmi Eta Kare Etki Büyüklüğüdür.
* sembolü ,05 düzeyinde, ** sembolü ise ,01 düzeyinde anlamlı *F* değerlerini göstermektedir. Tüm etkiler için geçerli olan hata serbestlik dereceleri şu şekildedir: 496 (Pillai İzi), 494 (Wilks Lambda), 492 (Hotelling İzi), 248 (Roy En Büyük Kökü).

Görüldüğü gibi iki yöntem arasında, istatistiksel anlamlılık değerleri açısından ciddi farklar olmamasına ve etki büyüklükleri genelde paralel değerlerden oluşmasına karşın, iki etkileşim etkisine ve bir temel etkiye ait etki büyükleri açısından ciddi farklar çıkmıştır. Şöyle ki, Ölçek Uzunluğu, Model x Ölçek Uzunluğu ve Ölçek

Uzunluğu x Kategori sayısı etkilerine ait etki büyüklükleri duble çok değişkenli modelde, çokdeğişkenli karma modelde olduğundan çok daha yüksektir. Test uzunluğuna ait etki büyüklükleri, aralarında fark olmasına karşın her iki modelde de yüksek olduğu için bu farkın sonuçların yorumlanmasında önemli bir sorun yaratmayacağı düşünülmüştür. Diğer iki etkinin ne derecede ciddiye alınması gerektiği ile ilgili yorumlarda ise ihtiyatlı davranılması gerekmektedir.

Birden fazla bağımlı değişkeni olan araştırma desenlerindeki müteakip analiz tercihlerinden birisi (bazı âlimlerin karşı çıkmasına rağmen) bağımsız değişkenlerin her bir bağımlı değişken üzerindeki etkilerinin ayrı ayrı incelenmesi yönündendir.[137] Bu çerçevede yapılan tek değişkenli analizlere dair sonuçlar Tablo 4.5.'de sunulmuştur.

Bu tip çalışmalarda çokdeğişkenlinin yanı sıra tekdeğişkenli analizin de yapılmasının gerekçesi; bağımsız değişkenlerin düzeyleri arasında farkın, aslında ağırlıklı olarak sadece bazı bağımlı değişkenler açısından mı geçerli olduğunu kontrol etme isteğidir. Bir başka deyişle düzeyler arasında farkın, en çok hangi bağımlı değişken üzerinde kendini hissettirdiğini tespit etmek amacıyla bu prosedür uygulanmaktadır. Sonuçlar incelendiğinde bağımsız değişkenlerin etkisinin Görgül Hata Varyansı değişkeni üzerinde çok daha güçlü olduğu görülmektedir. Yani ifade sayısını, kategori sayısını ve örneklem büyüklüğünü arttırmak genelde daha az hatalı tahminlemeler yapmamızı sağlamakla birlikte, asıl olarak tahminlemelerdeki istikrarın artmasına katkıda bulunmaktadır.

---

[137] Her ne kadar bazı yazarlar (mesela Weinfurt, 1995) çokdeğişkenli analizi takiben yapılan tek bağımlı değişkenli incelemeleri, müteakip analiz türleri arasında saysa da, müteakip mukayeseler [*post hoc comparisons*] tabiri aslında "istatistiksel olarak anlamlı" çıkan umumi [*omnibus*] testlerin ardından, bağımlı değişkenin hangi düzeyleri arasında farkın olduğunu belirlemek için yapılan karşılaştırmalar için kullanılmaktadır.

Tablo 4.5. Tek Değişkenli Testler

	Bağımlı Değişkenler		*KT*	*sd*	*F*	*p*	η_p^2
M	*Tahminleme Yanlılığı*	Düzeltilmemiş	0,082	1	2,297	,132	,018
		$\hat{\varepsilon}_{gg}$	0,082	1,000	2,297	,132	,018
		$\hat{\varepsilon}_{hf}$	0,082	1,000	2,297	,132	,018
		ε_{alt}	0,082	1,000	2,297	,132	,018
	Görgül Hata Varyansı	Düzeltilmemiş	0,013	1	1,801	,182	,014
		$\hat{\varepsilon}_{gg}$	0,013	1,000	1,801	,182	,014
		$\hat{\varepsilon}_{hf}$	0,013	1,000	1,801	,182	,014
		ε_{alt}	0,013	1,000	1,801	,182	,014
N	*Tahminleme Yanlılığı*	Düzeltilmemiş	0,009	1	10,211	,002	,076
		$\hat{\varepsilon}_{gg}$	0,009	1,000	10,211	,002	,076
		$\hat{\varepsilon}_{hf}$	0,009	1,000	10,211	,002	,076
		ε_{alt}	0,009	1,000	10,211	,002	,076
	Görgül Hata Varyansı	Düzeltilmemiş	0,003	1	21,499	<,001	,148
		$\hat{\varepsilon}_{gg}$	0,003	1,000	21,499	<,001	,148
		$\hat{\varepsilon}_{hf}$	0,003	1,000	21,499	<,001	,148
		ε_{alt}	0,003	1,000	21,499	<,001	,148
L	*Tahminleme Yanlılığı*	Düzeltilmemiş	1,008	2	10,916	<,001	,081
		$\hat{\varepsilon}_{gg}$	1,008	1,809	10,916	<,001	,081
		$\hat{\varepsilon}_{hf}$	1,008	1,834	10,916	<,001	,081
		ε_{alt}	1,008	1,000	10,916	<,001	,081
	Görgül Hata Varyansı	Düzeltilmemiş	0,786	2	104,666	<,001	,458
		$\hat{\varepsilon}_{gg}$	0,786	1,722	104,666	<,001	,458
		$\hat{\varepsilon}_{hf}$	0,786	1,744	104,666	<,001	,458
		ε_{alt}	0,786	1,000	104,666	<,001	,458
C	*Tahminleme Yanlılığı*	Düzeltilmemiş	0,493	1	14,173	<,001	,103
		$\hat{\varepsilon}_{gg}$	0,493	1,000	14,173	<,001	,103
		$\hat{\varepsilon}_{hf}$	0,493	1,000	14,173	<,001	,103
		ε_{alt}	0,493	1,000	14,173	<,001	,103
	Görgül Hata Varyansı	Düzeltilmemiş	1,530	1	118,141	<,001	,488
		$\hat{\varepsilon}_{gg}$	1,530	1,000	118,141	<,001	,488
		$\hat{\varepsilon}_{hf}$	1,530	1,000	118,141	<,001	,488
		ε_{alt}	1,530	1,000	118,141	<,001	,488

Tablo 4.5. (Devam).

	Bağımlı Değişkenler		*KT*	*sd*	*F*	*p*	η_p^2
M x N	*Tahminleme Yanlılığı*	Düzeltilmemiş	0,004	1	6,150	,014	,047
		$\hat{\varepsilon}_{gg}$	0,004	1,000	6,150	,014	,047
		$\hat{\varepsilon}_{hf}$	0,004	1,000	6,150	,014	,047
		ε_{alt}	0,004	1,000	6,150	,014	,047
	Görgül Hata Varyansı	Düzeltilmemiş	0,001	1	7,798	,006	,059
		$\hat{\varepsilon}_{gg}$	0,001	1,000	7,798	,006	,059
		$\hat{\varepsilon}_{hf}$	0,001	1,000	7,798	,006	,059
		ε_{alt}	0,001	1,000	7,798	,006	,059
M x L	*Tahminleme Yanlılığı*	Düzeltilmemiş	0,010	2	1,009	,366	,008
		$\hat{\varepsilon}_{gg}$	0,010	1,815	1,009	,360	,008
		$\hat{\varepsilon}_{hf}$	0,010	1,841	1,009	,361	,008
		ε_{alt}	0,010	1,000	1,009	,317	,008
	Görgül Hata Varyansı	Düzeltilmemiş	0,001	2	0,572	,565	,005
		$\hat{\varepsilon}_{gg}$	0,001	1,763	0,572	,544	,005
		$\hat{\varepsilon}_{hf}$	0,001	1,787	0,572	,546	,005
		ε_{alt}	0,001	1,000	0,572	,451	,005
N x L	*Tahminleme Yanlılığı*	Düzeltilmemiş	0,004	2	7,067	,001	,054
		$\hat{\varepsilon}_{gg}$	0,004	1,394	7,067	,004	,054
		$\hat{\varepsilon}_{hf}$	0,004	1,405	7,067	,004	,054
		ε_{alt}	0,004	1,000	7,067	,009	,054
	Görgül Hata Varyansı	Düzeltilmemiş	0,001	2	14,815	<,001	,107
		$\hat{\varepsilon}_{gg}$	0,001	1,512	14,815	<,001	,107
		$\hat{\varepsilon}_{hf}$	0,001	1,527	14,815	<,001	,107
		ε_{alt}	0,001	1,000	14,815	<,001	,107
M x N x L	*Tahminleme Yanlılığı*	Düzeltilmemiş	0,003	2	5,302	,006	,041
		$\hat{\varepsilon}_{gg}$	0,003	1,398	5,302	,013	,041
		$\hat{\varepsilon}_{hf}$	0,003	1,409	5,302	,013	,041
		ε_{alt}	0,003	1,000	5,302	,023	,041
	Görgül Hata Varyansı	Düzeltilmemiş	<0,001	2	4,447	,013	,035
		$\hat{\varepsilon}_{gg}$	<0,001	1,379	4,447	,025	,035
		$\hat{\varepsilon}_{hf}$	<0,001	1,389	4,447	,024	,035
		ε_{alt}	<0,001	1,000	4,447	,037	,035

Tablo 4.5. (Devam).

	Bağımlı Değişkenler		*KT*	*sd*	*F*	*p*	η_p^2
M x C	*Tahminleme Yanlılığı*	Düzeltilmemiş	<0,001	1	0,010	,919	<,001
		$\hat{\varepsilon}_{gg}$	<0,001	1,000	0,010	,919	<,001
		$\hat{\varepsilon}_{hf}$	<0,001	1,000	0,010	,919	<,001
		ε_{alt}	<0,001	1,000	0,010	,919	<,001
	Görgül Hata Varyansı	Düzeltilmemiş	0,003	1	0,995	,321	,008
		$\hat{\varepsilon}_{gg}$	0,003	1,000	0,995	,321	,008
		$\hat{\varepsilon}_{hf}$	0,003	1,000	0,995	,321	,008
		ε_{alt}	0,003	1,000	0,995	,321	,008
N x C	*Tahminleme Yanlılığı*	Düzeltilmemiş	<0,001	1	2,240	,137	,018
		$\hat{\varepsilon}_{gg}$	<0,001	1,000	2,240	,137	,018
		$\hat{\varepsilon}_{hf}$	<0,001	1,000	2,240	,137	,018
		ε_{alt}	<0,001	1,000	2,240	,137	,018
	Görgül Hata Varyansı	Düzeltilmemiş	<0,001	1	0,186	,667	,001
		$\hat{\varepsilon}_{gg}$	<0,001	1,000	0,186	,667	,001
		$\hat{\varepsilon}_{hf}$	<0,001	1,000	0,186	,667	,001
		ε_{alt}	<0,001	1,000	0,186	,667	,001
M x N x C	*Tahminleme Yanlılığı*	Düzeltilmemiş	<0,001	1	0,406	,525	,003
		$\hat{\varepsilon}_{gg}$	<0,001	1,000	0,406	,525	,003
		$\hat{\varepsilon}_{hf}$	<0,001	1,000	0,406	,525	,003
		ε_{alt}	<0,001	1,000	0,406	,525	,003
	Görgül Hata Varyansı	Düzeltilmemiş	<0,001	1	4,272	,041	,033
		$\hat{\varepsilon}_{gg}$	<0,001	1,000	4,272	,041	,033
		$\hat{\varepsilon}_{hf}$	<0,001	1,000	4,272	,041	,033
		ε_{alt}	<0,001	1,000	4,272	,041	,033
L x C	*Tahminleme Yanlılığı*	Düzeltilmemiş	0,013	2	1,149	,319	,009
		$\hat{\varepsilon}_{gg}$	0,013	1,507	1,149	,307	,009
		$\hat{\varepsilon}_{hf}$	0,013	1,522	1,149	,308	,009
		ε_{alt}	0,013	1,000	1,149	,286	,009
	Görgül Hata Varyansı	Düzeltilmemiş	0,048	2	8,835	<,001	,067
		$\hat{\varepsilon}_{gg}$	0,048	1,434	8,835	,001	,067
		$\hat{\varepsilon}_{hf}$	0,048	1,446	8,835	,001	,067
		ε_{alt}	0,048	1,000	8,835	,004	,067

Tablo 4.5. (Devam).

	Bağımlı Değişkenler		*KT*	*sd*	*F*	*p*	η_p^2
M x L x C	*Tahminleme Yanlılığı*	Düzeltilmemiş	0,005	2	3,009	,051	,024
		$\hat{\varepsilon}_{gg}$	0,005	1,834	3,009	,056	,024
		$\hat{\varepsilon}_{hf}$	0,005	1,860	3,009	,055	,024
		ε_{alt}	0,005	1,000	3,009	,085	,024
	Görgül Hata Varyansı	Düzeltilmemiş	0,001	2	1,608	,202	,013
		$\hat{\varepsilon}_{gg}$	0,001	1,723	1,608	,206	,013
		$\hat{\varepsilon}_{hf}$	0,001	1,745	1,608	,205	,013
		ε_{alt}	0,001	1,000	1,608	,207	,013
N x L x C	*Tahminleme Yanlılığı*	Düzeltilmemiş	<0,001	2	0,011	,989	<,001
		$\hat{\varepsilon}_{gg}$	<0,001	1,752	0,011	,982	<,001
		$\hat{\varepsilon}_{hf}$	<0,001	1,775	0,011	,983	<,001
		ε_{alt}	<0,001	1,000	0,011	,916	<,001
	Görgül Hata Varyansı	Düzeltilmemiş	<0,001	2	0,287	,751	,002
		$\hat{\varepsilon}_{gg}$	<0,001	1,350	0,287	,662	,002
		$\hat{\varepsilon}_{hf}$	<0,001	1,360	0,287	,664	,002
		ε_{alt}	<0,001	1,000	0,287	,593	,002
M x N x L x C	*Tahminleme Yanlılığı*	Düzeltilmemiş	<0,001	2	0,411	,663	,003
		$\hat{\varepsilon}_{gg}$	<0,001	1,876	0,411	,650	,003
		$\hat{\varepsilon}_{hf}$	<0,001	1,904	0,411	,653	,003
		ε_{alt}	<0,001	1,000	0,411	,523	,003
	Görgül Hata Varyansı	Düzeltilmemiş	<0,001	2	2,372	,095	,019
		$\hat{\varepsilon}_{gg}$	<0,001	1,265	2,372	,118	,019
		$\hat{\varepsilon}_{hf}$	<0,001	1,272	2,372	,118	,019
		ε_{alt}	<0,001	1,000	2,372	,126	,019

Not. KT: Kareler Toplamı, $\hat{\varepsilon}_{gg}$: Greenhouse-Geisser epsilon tahminlemesi, $\hat{\varepsilon}_{hf}$: Huynh-Feldt epsilon tahminlemesi, ε_{alt}: Düzeltilmemiş satırlarındaki değerler epsilon alt sınırına göre hesaplanmıştır. Hatalara ilişkin değerlere tabloda yer verilmemiştir.

Tablo 4.6. Test Uzunluğu ile İlgili Polinom Planlı Mukayeseleri

Etkiler	Bağımlı Değişken	Polinom	F	*p*	η_p^2
L	Tahminleme Yanlılığı	Doğrusal	19,324**	<,001	,135
		2. Dereceden	0,704	,403	,006
	Görgül Hata Varyansı	Doğrusal	165,478**	<,001	,572
		2. Dereceden	1,727	,191	,014
M x L	Tahminleme Yanlılığı	Doğrusal	1,563	,214	,012
		2. Dereceden	0,026	,871	<,001
	Görgül Hata Varyansı	Doğrusal	0,818	,368	,007
		2. Dereceden	0,106	,745	,001
N x L	Tahminleme Yanlılığı	Doğrusal	8,620*	,004	,065
		2. Dereceden	3,320	,071	,026
	Görgül Hata Varyansı	Doğrusal	22,390**	<,001	,153
		2. Dereceden	0,884	,349	,007
M x N x L	Tahminleme Yanlılığı	Doğrusal	5,863*	,017	,045
		2. Dereceden	3,695	,057	,029
	Görgül Hata Varyansı	Doğrusal	5,685*	,019	,044
		2. Dereceden	1,332	,251	,011
L x C	Tahminleme Yanlılığı	Doğrusal	1,234	,269	,010
		2. Dereceden	0,883	,349	,007
	Görgül Hata Varyansı	Doğrusal	11,600*	,001	,086
		2. Dereceden	0,164	,686	,001
M x L x C	Tahminleme Yanlılığı	Doğrusal	4,342*	,039	,034
		2. Dereceden	0,544	,462	,004
	Görgül Hata Varyansı	Doğrusal	2,130	,147	,017
		2. Dereceden	0,396	,530	,003
N x L x C	Tahminleme Yanlılığı	Doğrusal	0,003	,956	<,001
		2. Dereceden	0,026	,873	<,001
	Görgül Hata Varyansı	Doğrusal	0,383	,537	,003
		2. Dereceden	0,014	,906	<,001
M x N x L x C	Tahminleme Yanlılığı	Doğrusal	0,018	,894	<,001
		2. Dereceden	0,782	,378	,006
	Görgül Hata Varyansı	Doğrusal	3,112	,080	,024
		2. Dereceden	<0,001	,992	<,001

Not. Test uzunluğu üç düzeyden oluştuğu için sadece doğrusal ve 2. dereceden ilişkiler test edilmiştir.
* sembolü ,05 düzeyinde, ** sembolü ise ,01 düzeyinde anlamlı *F* değerlerini göstermektedir.

Mevcut araştırmada hiçbir birimler-arası değişkenin olmaması doğal olarak standart müteakip mukayese prosedürlerinin uygulanamamasına neden olmaktadır. Bununla beraber, bütünüyle denekler-içi tasarımlarda planlı mukayeselerin yapılmasına bir engel bulunmamaktadır.[138] Farklı alternatiflerin olmasına karşın bağımsız değişkenlerin[139] yapısı ve ölçüm düzeyleri ve temel araştırma soruları dikkate alındığında en uygun prosedürün polinom trend (eğilimler) analizi olduğu düşünülmüştür (bkz. Tablo 4.6.).

Burada merak edilen ilişki, test uzunluğu ile bağımlı değişkenler arasındaki ilişinin biçimidir ki bunun her iki bağımlı değişken için de doğrusal olma ihtimalinin eğrisele göre çok daha yüksek olduğu tespit edilmiştir.

5. SONUÇ, GENEL DEĞERLENDİRME VE İLERİKİ ÇALIŞMALAR İÇİN ÖNERİLER

Çalışma çerçevesinde araştırma sorularına bağlı olarak geliştirilen tüm maddi hipotezleri destekler nitelikte bulgular elde edilmiştir. Bu sonuçlardan en etkileyici olanı Çok İhtimalli Rasch Modeli ile Derecelendirilmiş Yanıt Modeli arasında kişi parametreleri tahminleme performansı açısından fark olmadığı sonucudur. Bu bulgu, sadece pratik kaygılar taşıyan bir araştırmacının bile çok ihtimalli ölçekler kullanırken 2 PL modellere yönelmesini gerektirecek bir durumun bulunmadığı anlamına gelmektedir. Bir başka değişle, orta büyüklükte örneklemlerle ve kısa anketlerle çalışan örgütsel davranış araştırmacıları için Rasch Modelleri her bakımdan en iyi seçeneği oluşturmaktadır.

Soru sayısı, örneklem büyüklüğü, şık sayısının artmasıyla tahminleme performansın da artacağı önceden de bilinen bir gerçektir. Dolayısıyla bu hipotezleri

---

[138] Ancak planlı müteakip analizler, yalnızca bağımlı değişkenlerin her biri ayrı ayrı ele alınıp uygulanabilmektedir. Bu analizlerin çok değişkenli versiyonlarının da olduğuna dair yazarın bir malumatı bulunmamaktadır.

[139] Aslında iki ihtimalli ifadeleri modelden çıkardığımız için düzey sayısı (k) ikiden büyük sadece bir bağımlı değişkenimiz kalmıştır ki o da ölçek uzunluğudur. Dolayısıyla söz konusu planlı mukayeseler sadece ölçek uzunluğu ile ilgili temel ve etkileşim etkilerine uygulanmıştır.

destekler nitelikte sonuçların elde edilmesinin bir etkileyiciliği bulunmamaktadır. Bununla beraber ölçeği kısa tutup, şık sayısını arttırmak ile şık sayısını düşük tutup daha uzun bir ölçek kullanma arasında bir tercih yapılması gerektiğinde çok kategorili az sorulu ölçeklerin diğer alternatife göre daha iyi sonuçlar verdiği ortaya çıkmıştır ki bu nispeten orijinal bir bilgidir.

ÇİRM'nin örneklem büyüklüğünün azalmasında hemen hemen hiç etkilenmediğini ancak DYM'nin örneklem büyüklüğündeki azalıştan biraz etkilendiği sonucuna varılmıştır.

İki bağımlı değişken açsından ayrı ayrı incelendiğinde, şık sayısı artışı ile tahminleme başarısı arasında eğrisel değil doğrusal bir ilişkinin olduğu sonucu bulunmuştur.

Kategori Sayısı x Ölçek Uzunluğu etkileşimi dışında faktörler arasında önemli bir etkileşim etkisinin olmadığı tespit edilmiş, ancak Kategori Sayısı x Test Uzunluğu Etkileşimin gücünün modelden modele farklılık gösterdiği bulunmuştur.

Simülasyon çalışmalarında mantıklı olabileceği düşünülen tüm faktörlerin modele dâhil edilmemesinin ve incelenen her faktöre ait düzeylerin çok ayrıntılı olarak belirlenmemesinin en önemli nedeni ilave edilen her faktörün ve düzeyin veri üretimi ve analizlerle ilgili iş yükünü katlayarak arttırmasıdır. Bu aynı zamanda yineleme sayısının birçok çalışmada çok düşük tutulmasının da temel nedenidir. Dolayısıyla araştırmacılar, modellerini kurarken faktör, düzey ve yineleme sayısı açısından ekonomik olmak zorunda kalırlar (yalnız bir araştırmayla, konuyla ilgili bütün sorulara cevap bulunamayacağı için bu gayet normal bir durumdur). Bu çalışmayı takip edecek çalışmalarda araştırmacılar, evvelki çalışmalarının sonuçlarına göre hangi faktörler üzerinde daha ayrıntılı durmaları gerektiğine, hangi faktörlerin modelden çıkarılıp yerine başka faktörlerin koyulabileceğine karar verirler. Bu çerçevede ileriki çalışmalar için şu gibi önerilerde bulunulabilir:

(a) Bu çalışmada örgütsel davranış bilimi açısından en tipik durumlarda hangi modelin daha iyi sonuçlar verdiği tespit edilmeye çalışılmıştır. İleriki çalışmalarda örneklem büyüklüğü ve soru sayısı açısından modellerin başabaş noktası tespit edilmeye çalışılabilinir. Örneklem büyüklüğü arttıkça 2 PL modellerin tahminleme

performanslarının arttığı bilinmektedir.[140] 2 PL modelin Rasch Modelinden daha iyi sonuçlar vermeye başladığı örneklem büyüklüğü noktası tespit edilmeye çalışılabilir. Bunun için kalibrasyon örneklemi büyüklüğü faktörü için 6-7 eşit aralıklı düzey belirlenerek araştırma tekrarlanabilir.

(b) Benzer şekilde soru sayısı açısından da iki modelin başa baş noktası incelenebilir.

(c) Kategori sayısındaki artışın da tahminleme performansını arttırdığı bilinmesine karşın tipik uygulamalar 4 veya 5 şıklı ifadelerden oluştuğundan bu faktör ileriki çalışmalarda model dışında (mesela dörde sabitlenerek) tutulabilir.

(d) Bu araştırmada sabit tutulan ayırt edicilik parametresinin dağılımına ilişkin çeşitli kombinasyonlar denenebilir. Mesela bu çalışmada 0,8 ile 2,0 arasında dikdörtgensel dağılım gösteren ayırt edicilik parametreleri incelenmişken ileriki çalışmalarda ortalaması 1 olan, normal ve sola çarpık dağılım gösteren parametrelerle çalışılması ilginç sonuçların elde edilmesini sağlayabilir.

(e) Her ne kadar MYK simülasyon araştırmalarında yineleme sayısı genelde düşük tutulmasına ve bunun bir çok araştırmacı tarafından büyük bir kusur olarak görülmemesine rağmen yineleme sayısı dikkatle ele alınması gereken bir konudur ve bu çalışmanın en büyük eksikliği yineleme sayısının oldukça düşük olmasıdır. MYK çerçevesindeki, hiç yineleme yapmayan onlarca çalışmaya göre bu çalışmada için belirlenen yineleme[141] sayısı (5) düşük olarak kabul edilmeyebilir ancak, mesela yapısal eşitlik modellemesi araştırmalarındaki yineleme rakamlarıyla karşılaştırıldığında (ayrıntılar için bkz. Bandalos, 2006) bu sayı çok düşük kalmaktadır. İleriki çalışmalarda yineleme sayısı, simülasyon çalışmasına daha müsait alternatif yazılımların kullanılmasıyla arttırılabilir.

---

[140] Bu Rasch Modeli için de geçerlidir ama 2 PL deki performans artışı Rasch Modelinkinden daha hızlıdır.

[141] Bu çalışmada bu sayının düşük tutulmasının en büyük nedeni Multilog programının simülasyon çalışmalarına müsait olmayan bir yapısının olması ve Winsteps ile yapılan analizlerin aksine bu yazılım ile işlemlerin otomatik olarak yaptırılamamasıdır. Bu da çalışma için gereken iş yükünü çok ciddi şekilde arttırmıştır.

EKLER

Ek 1: Tezde Kullanılan Bazı Kavramların İngilizce Orijinalleri

Alt Asimptot	Lower Asymptote
Berklik	Robustness
Birim	Subject
Birimler-arası	Between-subjects
Birimler-içi	Within-subjects
Çeldiriciler	Item Distracters
Çokdeğişkenli - Çok Bağımlı Değişkenli	Multivariate
Çok İhtimalli	Polytomous
Denek	Respondent- Subject
Dereceleme Ölçeği Modeli	Rating Scale Model
Derecelendirilmiş Yanıt Modeli	Graded Response Model
Duble Çokdeğişkenli Varyans Analizi (DÇVA)	Doubly Multivariate ANOVA
Farksızlık Hipotezi İstatistiksel Anlamlılık Testi	Null Hypothesis Statistical Significance Testing
İki İhtimalli	Dichotomous
Karma Tasarım	Mixed Design
Madde Karakteristiği Eğrisi (MKE)	Item Characteristic Curve (ICC)
Madde Yanıt Kuramı (MYK)	Item Response Theory (IRT)
Mukayesebaşı Hata Oranı	Per-comparison Error Rate
Normal Kemer Modeli	Normal Ogive Model
Operasyonel - işlemsel - işevuruk Tanım	Operational Definition
Etkiboyu Hata Oranı	Familywise Error Rate
Deneyboyu Hata Oranı	Experimentwise Error Rate
Tahminleme	Estimation

Kısmi Puanlama Modeli	Partial Credit Model
Olabilirlik	Likelihood
Ayrıştırılabilirlik	Separatebility
Müşterek Ençok Olabilirlik	Joint Maximum Likelihood
Maddi Hipotez	Substantive Hypothesis
Sanal Denek	Simulee
Sabit Etkiler Modeli	Fixed Effect Model
Devam Eden Tesir	Carry-over Effect
Eşlik Değişkeni	Covariate
Müteakip Mukayeseler	Post Hoc Comparison
Küresellik	Sphericity
Dışadüşen	Outlier
Kareleme	Quadrature
Ölçek/birim Denkleştirme Faktörü	Scaling Factor
Örtük Özellik Kuramı (ÖÖK)	Latent Trait Theory
Umumî (tümel) Test	Ominibus test
Uzanım	Continuum
Üst Asimptot	Upper Asymptote
Yenileme-Yenilemeli	Iteration-Iterative
Yineleme	Replication

Ek 2: Farksızlık Hipotezi İstatistiksel Anlamlılık Testleri (FYİAT) ile İlgili Açıklama

Anlamlılık testleri, sosyal bilimlerde çok yaygın olarak kullanılan bir analiz yöntemidir. Ancak anlamlılık testlerinin kullanılmasına yönelik ciddi eleştiriler de yapılmaktadır. Bu eleştirilerin ilk çıkışı 1940'lara kadar gitmektedir. Anlamlılık testlerinin, aynı anda bu kadar uzun zamandır ve yoğun bir şekilde eleştirilirken bir yandan da bu kadar yaygın bir şekilde kullanılması oldukça ilginçtir. Daha da ilginci, anlamlılık testlerini eleştirenlere (bugüne kadar anlamlılık testlerini eleştiren 200'ün üzerinde makale yazılmış ve birçok bilimsel dergide bu konuya yönelik özel sayılar düzenlemiştir) karşı, bu yaklaşımı savunanların oluşturduğu güçlü bir karşı kampın da bulunmamasıdır. Bu durumda sosyal bilimcilerin çoğunun (anlamlılık testlerinin sakıncalarından haberdar olmayan kesim bir kenara bırakılırsa) eleştirilere katılmakla birlikte; alışkanlık, işin kolayına kaçma gibi nedenlerle bu sakıncalı uygulamayı sürdürdükleri sonucuna varılabilir.

FYİAT uygulamalarına karşı çıkan âlimler, karşı olma derecelerine göre farklı gruplara ayrılabilir. Mesela, Cohen (1994) ve Gigerenzer (1993) en uç düzeyde FYİAT uygulamalarını eleştirirken, Kline (2004) bazı istisnai durumlarda anlamlılık testlerinin bir kullanım alanı olabileceğini ancak birçok çalışma için bu uygulamayı tercih etmenin büyük bir hata olduğunu belirtmiştir. Goodman (1993), Hubbard ve Armstrong (2006) gibi bazı yazarlar, FYİAT ile ilgili sorunların nedeni olarak günümüzde hipotez testi adıyla izlenen prosedürlerin aslında bir biriyle uyumlu olmayan iki test yaklaşımının (Fisher'in anlamlılık testi ve Neyman-Pearson'ın Hipotez testi) karışımı, melez bir yol olmasını görürken, Thompson (2002) gibi bazı yazarlar ise FYİAT'ın kullanımını değil, etki büyüklüğü, güven aralığı gibi yöntemlere araştırmalarda yeterince yer verilmemesini eleştirmektedir. Amerikan Psikoloji Derneği (APA) ise anlamlılık testlerine ait bilgilerin makalelerde sunulmasına müdahale etmezken, sonuca ilişkin kararların sadece bu testlere göre yapılmaması gerektiğini, etki büyüklüğü, güven aralığı veya etki büyüklüklerine ilişkin güven aralılarına ait bilgileri sunmayan çalışmaların eksik olarak görüleceğini belirtmiştir (2009, s. 33).

Bu çalışmada, aslında Cohen'inkine yakın bir görüşe sahip olunmasına rağmen, APA'nın yönlendirmeleri doğrultunda (bu istatistiklere aşina okuyucuların daha rahat

edeceğini de düşünerek) hareket edilmiş, yorumlar etki büyüklükleri üzerinden yapılırken anlamlılık değerleri de metin içinde ve tablolarda sunulmaya devam edilmiştir. APA'nın önermesine (s.34) karşın etki büyüklüklerine ilişkin güven aralıkları, bunları (çok değişkenli ve tekrarlı ölçümler içeren modellerin etki büyüklüklerinin güven aralıklarını) hesaplayabilen yazılımlar henüz geliştirilmediği için, raporlanamamıştır. Bununla beraber, tez içersinde "hipotez testi", "sonuçların anlamlılığı" gibi tabirlerin kullanılmasından mümkün olduğunca kaçınılmaya çalışılmıştır.

EK 3: Tablolar

Tablo 1. Tahminleme Yanlılığı için Model x Kalibrasyon Örneklem Büyüklüğü x Ölçek Uzunluğu x Kategori Sayısı İstatistikleri

Model	Kalibrasyon Örneklem Büyüklüğü	Ölçek Uzunluğu	Kategori Sayısı	Ortalama	Standart Hata	% 95 Güven Aralığı Alt Sınır	Üst Sınır
ÇİRM	300	12	6	,141	,010	,122	,160
			4	,159	,012	,135	,182
		9	6	,153	,010	,133	,173
			4	,181	,013	,156	,206
		6	6	,182	,013	,157	,206
			4	,215	,015	,184	,245
	150	12	6	,142	,010	,123	,161
			4	,159	,012	,136	,182
		9	6	,155	,010	,135	,175
			4	,182	,013	,157	,206
		6	6	,183	,013	,158	,208
			4	,216	,016	,185	,247
DYM	300	12	6	,152	,011	,131	,173
			4	,175	,012	,152	,198
		9	6	,162	,011	,141	,183
			4	,191	,012	,167	,216
		6	6	,186	,013	,161	,211
			4	,212	,015	,182	,242
	150	12	6	,155	,011	,134	,177
			4	,177	,012	,154	,200
		9	6	,166	,011	,144	,188
			4	,195	,013	,170	,219
		6	6	,198	,013	,173	,224
			4	,223	,015	,192	,253

Not. ÇİRM: Çok İhtimalli Rasch Modeli, DYM: Derecelendirilmiş Yanıt Modeli

Tablo 2. Mauchly Küresellik Testi*

Birimler-içi Etkiler	Bağımlı Değişkenler	*Mauchly W*	*p*	$\hat{\varepsilon}_{gg}$	$\hat{\varepsilon}_{hf}$
L	Tahminleme Yanlılığı	,894	,001	,904	,917
	Görgül Hata Varyansı	,838	,<,001	,861	,872
M * L	Tahminleme Yanlılığı	,898	,001	,908	,920
	Görgül Hata Varyansı	,866	,000	,882	,894
N * L	Tahminleme Yanlılığı	,565	<,001	,697	,703
	Görgül Hata Varyansı	,677	<,001	,756	,763
M * N * L	Tahminleme Yanlılığı	,569	<,001	,699	,704
	Görgül Hata Varyansı	,550	<,001	,690	,695
L * C	Tahminleme Yanlılığı	,673	<,001	,754	,761
	Görgül Hata Varyansı	,605	<,001	,717	,723
M * L * C	Tahminleme Yanlılığı	,909	,003	,917	,930
	Görgül Hata Varyansı	,839	<,001	,861	,872
N * L * C	Tahminleme Yanlılığı	,858	<,001	,876	,887
	Görgül Hata Varyansı	,519	<,001	,675	,680
M * N * L * C	Tahminleme Yanlılığı	,934	,015	,938	,952
	Görgül Hata Varyansı	,419	<,001	,632	,636

Not. $\hat{\varepsilon}_{gg}$: Greenhouse-Geisser epsilon tahminlemesi, $\hat{\varepsilon}_{hf}$: Huynh-Feldt epsilon tahminlemesi, M: Model, N: Kalibrasyon Örneklemi Büyüklüğü, L: Ölçek Uzunluğu, C: Kategori Sayısı, *p*: *Mauchly W* değerinin Ki Kare dağılımına yaklaştırılması ile elde edilen İstatistiğinin Anlamlılık Seviyesidir.

* $\varepsilon_{alt} = {}^{1}/_{k-1}$, olduğundan ki burada *k* düzey sayısını temsil etmektedir, tüm etkiler için epsilon tahminlemesi alt sınırı (ε_{alt}) ,05'dir.

Tablo 3. Görgül Hata Varyansı için Model x Kalibrasyon Örneklem Büyüklüğü x Ölçek Uzunluğu x Kategori Sayısı İstatistikleri

Model	Kalibrasyon Örneklem Büyüklüğü	Ölçek Uzunluğu	Kategori Sayısı	Ortalama	Standart Hata	% 95 Güven Aralığı Alt Sınır	% 95 Güven Aralığı Üst Sınır
ÇİRM	300	12	6	,065	,003	,059	,071
			4	,105	,006	,092	,118
		9	6	,079	,004	,071	,086
			4	,125	,008	,110	,141
		6	6	,096	,005	,086	,107
			4	,154	,008	,138	,169
	150	12	6	,066	,003	,060	,072
			4	,105	,006	,092	,117
		9	6	,080	,004	,072	,088
			4	,125	,008	,110	,141
		6	6	,099	,005	,088	,110
			4	,154	,008	,138	,170
DYM	300	12	6	,074	,003	,068	,079
			4	,106	,004	,098	,114
		9	6	,085	,003	,079	,091
			4	,127	,005	,117	,137
		6	6	,099	,004	,092	,106
			4	,152	,006	,141	,164
	150	12	6	,075	,003	,069	,080
			4	,107	,004	,099	,115
		9	6	,086	,003	,080	,092
			4	,130	,005	,120	,140
		6	6	,103	,004	,096	,111
			4	,159	,006	,148	,170

Not. ÇİRM: Çok İhtimalli Rasch Modeli, DYM: Derecelendirilmiş Yanıt Modeli

Tablo 4. Çarpıklık Basıklık Katsayıları

İstatistikler

	N	Çarpıklık		Basıklık	
		İstatistik	Std. Hata	İstatistik	Std. Hata
Yanlılık-PN3L12C6	125	,655	,217	-,535	,430
Yanlılık-PN3L12C4	125	1,066	,217	,943	,430
Yanlılık-PN3L12C2	125	1,053	,217	,391	,430
Yanlılık-PN3L09C6	125	,779	,217	-,100	,430
Yanlılık-PN3L09C4	125	,899	,217	,646	,430
Yanlılık-PN3L09C2	125	,915	,217	,161	,430
Yanlılık-PN3L06C6	125	1,265	,217	1,682	,430
Yanlılık-PN3L06C4	125	1,293	,217	1,886	,430
Yanlılık-PN3L06C2	125	,964	,217	,757	,430
Yanlılık-PN1L12C6	125	,696	,217	-,462	,430
Yanlılık-PN1L12C4	125	1,046	,217	,863	,430
Yanlılık-PN1L12C2	125	1,075	,217	,454	,430
Yanlılık-PN1L09C6	125	,777	,217	-,136	,430
Yanlılık-PN1L09C4	125	,889	,217	,597	,430
Yanlılık-PN1L09C2	125	,928	,217	,209	,430
Yanlılık-PN1L06C6	125	1,221	,217	1,551	,430
Yanlılık-PN1L06C4	125	1,271	,217	1,803	,430
Yanlılık-PN1L06C2	125	,963	,217	,734	,430
Yanlılık-GN3L12C6	125	,981	,217	,647	,430
Yanlılık-GN3L12C4	125	,921	,217	,332	,430
Yanlılık-GN3L12C2	125	1,004	,217	,216	,430
Yanlılık-GN3L09C6	125	,860	,217	,400	,430
Yanlılık-GN3L09C4	125	,568	,217	-,577	,430
Yanlılık-GN3L09C2	125	,854	,217	,066	,430
Yanlılık-GN3L06C6	125	1,205	,217	1,641	,430
Yanlılık-GN3L06C4	125	1,194	,217	1,785	,430
Yanlılık-GN3L06C2	125	,875	,217	,415	,430
Yanlılık-GN1L12C6	125	1,034	,217	,979	,430
Yanlılık-GN1L12C4	125	,859	,217	,201	,430
Yanlılık-GN1L12C2	125	1,035	,217	,324	,430
Yanlılık-GN1L09C6	125	,816	,217	,441	,430
Yanlılık-GN1L09C4	125	,469	,217	-,693	,430
Yanlılık-GN1L09C2	125	,780	,217	-,058	,430
Yanlılık-GN1L06C6	125	1,092	,217	1,419	,430
Yanlılık-GN1L06C4	125	1,139	,217	1,754	,430
Yanlılık-GN1L06C2	125	,866	,217	,457	,430
Gör. Var.-PN3L12C6	125	2,366	,217	11,531	,430
Gör. Var.-PN3L12C4	125	1,502	,217	2,194	,430
Gör. Var.-PN3L12C2	125	,772	,217	,105	,430
Gör. Var.-PN3L09C6	125	1,802	,217	3,956	,430
Gör. Var.-PN3L09C4	125	1,357	,217	1,226	,430
Gör. Var.-PN3L09C2	125	,167	,217	-,461	,430
Gör. Var.-PN3L06C6	125	2,344	,217	7,476	,430
Gör. Var.-PN3L06C4	125	,916	,217	,172	,430

Tablo 4. Devam.

İstatistikler

	N	Çarpıklık		Basıklık	
		İstatistik	Std. Hata	İstatistik	Std. Hata
Gör. Var.-PN3L06C2	125	,397	,217	-,522	,430
Gör. Var.-PN1L12C6	125	2,477	,217	12,580	,430
Gör. Var.-PN1L12C4	125	1,511	,217	2,312	,430
Gör. Var.-PN1L12C2	125	,779	,217	,155	,430
Gör. Var.-PN1L09C6	125	1,830	,217	4,053	,430
Gör. Var.-PN1L09C4	125	1,361	,217	1,284	,430
Gör. Var.-PN1L09C2	125	,206	,217	-,485	,430
Gör. Var.-PN1L06C6	125	2,219	,217	6,486	,430
Gör. Var.-PN1L06C4	125	,902	,217	,259	,430
Gör. Var.-PN1L06C2	125	,432	,217	-,339	,430
Gör. Var.-GN3L12C6	125	1,087	,217	3,168	,430
Gör. Var.-GN3L12C4	125	,660	,217	,172	,430
Gör. Var.-GN3L12C2	125	,692	,217	-,078	,430
Gör. Var.-GN3L09C6	125	,517	,217	-,082	,430
Gör. Var.-GN3L09C4	125	,739	,217	-,027	,430
Gör. Var.-GN3L09C2	125	,414	,217	-,365	,430
Gör. Var.-GN3L06C6	125	,966	,217	1,305	,430
Gör. Var.-GN3L06C4	125	,295	,217	-,554	,430
Gör. Var.-GN3L06C2	125	,442	,217	-,481	,430
Gör. Var.-GN1L12C6	125	1,013	,217	2,690	,430
Gör. Var.-GN1L12C4	125	,728	,217	,220	,430
Gör. Var.-GN1L12C2	125	,573	,217	-,129	,430
Gör. Var.-GN1L09C6	125	,449	,217	,044	,430
Gör. Var.-GN1L09C4	125	,724	,217	,097	,430
Gör. Var.-GN1L09C2	125	,361	,217	-,329	,430
Gör. Var.-GN1L06C6	125	,799	,217	,826	,430
Gör. Var.-GN1L06C4	125	,335	,217	-,545	,430
Gör. Var.-GN1L06C2	125	,368	,217	-,373	,430
Valid N (listwise)	125				

KAYNAKÇA

Andrich, D. (1978). Application of a Psychometric Model to Order Categories Which are Scored with Successive Integers. *Applied Psychological Measurement*, 2, 581-594.

Andrich, D. (1982). An Index of Person Separation in Latent Trait Theory, the Traditional KR 20 Index and the Guttman Scale Response Pattern. *Educational Research and Perspectives*, 9, 95-104.

Andrich, D. (1988*). Rasch Models for Measurement* (Sage University Paper Series on Quantitative Applications in the Social Sciences). Beverly Hills: Sage Publications, Inc.

Andrich, D. (2002). Understanding Resistance to the Data-Model Relationship in Rasch's Paradigm: A Reflection for the Next Generation. *Journal of Applied Measurement*, 3/3, 325-329.

Andrich, D. (2004). Controversy and the Rasch Model: A Characteristic of Incompatible Paradigms? *Medical Care*, 24/1, 7-16.

Andrich, D. (2007). Rating Scale Analysis. Der. G. N. Masters ve J. P. Keeves, *Advances in Measurement in Educational Research and Assessment* (pp. 110-121). Bingley: Emerald Group Publishing Limited.

Andrich, D. (2008a). Introductory Course on Rasch Measurement. *Eğitim Notu.* The University of Western Australia.

Andrich, D. (2008b). Advanced Course on Rasch Measurement. *Eğitim Notu.* The University of Western Australia.

Andrich, D., de Jong, J. ve Sheridon, B. S. (1997). Diagnostic Opportunities with the Rasch Model for Ordered Response Categories. Der. J. Rost ve R. Langeheine, *Application of Latent Trait and Latent Class Model in the Social Sciences* (pp. 58-68). Munster, Germany: Waxmann.

Baghaei, P. (2008). Local Dependency and Rasch Measures. *Rasch Measurement Transactions*, 21/3, 1105-1106.

Baker, F. B. (2001). *The Basic of Item Response Theory* (2.bs.). ERIC Clearinghouse on Assessment and Evaluation.

Baker, F. B. ve Kim, S. H. (2004). *Item Response Theory: Parameter Estimatşon Techniques* (2. bs.). New York: Marcel Dekker, Inc.

Bandalos, D. L. (2006). The Use of Monte Carlo Studies in Structural Equation Modeling Research. (Der.) G. R. Handcock, ve R. O. Mueller. *Structural Equation Modeling* (Quantitative Methods in Education and the Behavioral Sciences: Issues, Research and Teaching). Connecticut: Information Age Publishing.

Bartholomew, D. J., Steele, F., Moustaki, I. ve Galbraith, J. I. (2002). *The Analysis and Interpretation of Multivariate Data for Social Scientists.* Boca Raton, Florida: CRC Press

Barton, M .A. ve Lord, F. M. (1981). *An Upper Asymptote for the Three-Parameter Logistic Item Response Model* (Tech. Rep. RR-81-20). Princeton, NJ: Educational Testing Service.

Baysal, A. C. (1981). *Sosyal ve Örgütsel Psikolojide Tutumlar.* İstanbul: İ.Ü. İşletme Fakültesi.

Birnbaum , A. (1968), Some Latent Trait Models and Their Use in Inferring an Examinee's Ability (Der.), F.M. Lord and M.R. Novick, *Statistical Theories of Mental Test Scores* (pp. 397-472). Reading, MA: Addison-Wesley.

Bock, R. D. ve Aitkin, M. (1981). Marginal Maximum Likelihood Estimation for the Item Parameters: Application of an E-M Algorithm, *Psychometrika*, 46/4, 443-459.

Boik, R. J. (1988). The Mixed Model for Multivariate Repeated Measures: Validity Conditions and an Approximate Test. *Psychometrika*, 53/4, 469-486.

Borsboom, D. (2005). *Measuring the Mind: Conceptual Issues in Contemporary Psychometrics.* Cambridge: Cambridge University Press.

Byrne, B. M. (2001). *Structural Equation Modeling with AMOS: Basic Concepts, Applications and Programming.* New Jersey: Lawrence Erlbaum Association, Inc.

Choi, S. W., Cook, K. F. ve Dodd, B. G. (1997). Parameter Recovery for the Partial Credit Model Using MULTILOG. *Journal of Outcome Measurement*, 1/2, 114-142.

Cohen, J. (1994). The Earth is round (p=.05). *American Psychologist*, 49, 997-1003.

Crocker, L. & Algina, J. (2006). *Introduction to Classical and Modern Test Theory*, Wadsworth. Orlando FL: Harcourt Brace.

de Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory.* New York: Guilford Press.

de Ayala, R. J. ve Bolesta, M. S. (1999). Item Parameter Recovery for the Nominal Response Model. *Applied Psychological Measurement*, 23, 3-20.

De Gruijter, D. N. M. ve van der Kamp, L. J. Th. (2008). *Statistical Test Theory for the Behavioral Sciences.* London: Chapman and Hall/CRC.

DuBois, B. ve Burns, J. A. (1975). An Analysis of the Meaning of the Question Mark Response Category in Attitude Scales. *Educational & Psychological Measurement*, 35, 869–884.

Durlak, J. A. (1995). Understanding Meta-Analysis. Der. L. G. Grimm ve P. R. Yarnold, *Reading and Understanding More Multivariate Statistics* (pp. 319-352). Washington DC: American Psychological Association.

Fitz-Gibbon, C. T. (1984). Meta Analysis: An Explication. *British Educational Research Journal*, 10, 135-144.

Fraser, C. ve McDonald, R. P. (2003). *NOHARM: A Windows Program for Fitting Both Unidimensional and Multidimensional Normal Ogive Models of Latent*

Trait Theory (Bilgisayar Programı). Welland, ON: Niagara College.

French, G. A. ve Dodd, B. G. (1999). Parameter Recovery for the Rating Scale Model Using PARSCALE. *Journal of Outcome Measurement*, 3/2, 176-199.

Gamst, G., Meyers, L. S. ve Guarino, A. J. (2008). *Analysis of Variance Designs: A Conceptual and Computational Approach with SPSS and SAS*. New York: Cambridge University Press.

Gigerenzer, G. (2009). The Superego, the Ego, and the Id in Statistical Reasoning. Der. G. Keren ve C. Lewis, *A Handbook for Data Analysis in the Behavioral Sciences: Methodological Issues* (pp. 311-335). New York: Psychology Press.

Glass, G. V. ve Hopkins, K. D. (1995), *Statistical Methods in Education and Psychology* (3. bs.). Boston: Allyn and Bacon.

Goodman, S. N. (1993). P Values, Hypothesis Tests, and Likelihood: Implication for Epidemiology of a Neglected Historical Debate. *American Journal of Epidemiology*, 137/5, 485-496.

Green, S. B., Salkind, N. J. ve Akey, T. M. (1997). *Using SPSS for Windows: Analyzing and Understanding Data* (2. bs.). New Jersey: Prentice-Hall, Inc.

Hambleton, R. K. ve Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Norwell, MA:Kluwer Academic Publisher.

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

Han, K. T. (2007). WinGen2: Windows software that generates IRT parameters and item responses *(Bilgisayar Programı)*. Amherst, MA: University of Massachusetts Amherst, Center for Educational Assessment.

Han, K. T. ve Hambleton, R. K. (2007). User's Manual: *WinGen* (*Center for Educational Assessment Report No. 642)*. Amherst, MA: University of Massachusetts, School of Education.

Handcock, G. R. ve Mueller, R. O. (2006). *Structural Equation Modeling* (Quantitative Methods in Education and the Behavioral Sciences: Issues, Research and Teaching). Connecticut: Information Age Publishing.

Harlow, L. L., Mulaik, S. A. ve Steiger, J. H. (Der.) (1997). *What If There were No Significance Tests?* New York: Psychology Press.

Harwell, M., Stone, C. A., Hsu, T.ve Kirisci, L. (1996). Monte Carlo Studies in Item Response Theory, *Applied Psychological Measurement*, 20, 101-125.

Henard, D. H. (2000). Item Response Theory. Der. L.G. Grimm ve P.R. Yarnold, *Reading and Understanding More Multivariate Statistics* (pp. 67-97). Washington DC: American Psychological Association.

Hubbard, R. ve Armstrong, J. S. (2006). Why We Don't Really Know What Statistical Significance Means: Implications for Educators. *Journal of Marketing Education*, 28, 114-120.

Kenny, D. A. (2003). "Unit of Analysis", (çevrimiçi)

http://davidakenny.net/u_o_a.htm, 15 Eylül 2009.

Keren, G. (1993). Between- or Within-Subjects Design: A Methodological Dilemma. Der. G. Keren ve C. Lewis, *A Handbook for Data Analysis in the Behavioral Sciences: Methodological Issues* (pp. 257-272). New York: Psychology Press

Kerlinger, F. N. ve Lee, H. B. (2000). *Foundation of Behavioral Research* (4. bs.). Orlando: Harcourt College Publishers.

Kline, R. B. (2004). *Beyond Significance Testing: Reforming Data Analysis Methods in Behavioral Research* (2.bs.). Washington: American Psychological Association.

Kline, R. B. (2009). *Becoming a Behavioral Science Researcher: A Guide to Producing Research That Matters.* New York: The Guilford Press.

Krech, D., Crutchfield, R. S. ve Ballachey, E. L. (1971). *Cemiyet İçinde Fert*, Çev. Mümtaz Turhan. İstanbul: Sosyal Bilimler Komisyonu Yayınları.

Kulas, J. T., Stachowski, A. A. ve Haynes, B. A. (2008). Middle Response Functioning in Likert Responses to Personality Items. *Journal of Business and Psychology, 22,* 251-260.

Licht, M. H. (1995). Multiple Regression and Correlation. Der. L. G. Grimm ve P. R. Yarnold, *Reading and Understanding More Multivariate Statistics* (pp. 19-64). Washington DC: American Psychological Association.

Likert, R. (1932). A Technique for the Measurement of Attitudes. *Archives of Psychology*, 140, 1–55.

Linacre, J. M. (1999). Understanding Rasch Measurement: Estimation Methods for Rasch Measures. *Journal of Outcome Measurement*, 3/4, 382-405.

Linacre, J. M. (2004a). Item Discrimination, Guessing and Carelessness: Estimating IRT Parameters with Rasch. *Rasch Measurement Transaction*, 18/1, 959-960.

Linacre, J. M. (2004b). Rasch Model Estimation: Further Topics. *Journal of Applied Measurement,* 5/1, 95-110.

Linacre, J. M. (2009). *A Users Guide to Winsteps / Ministep Rasch Model Computer Programs.*

Luchins, A. S. ve Luchins, E. H. (1950). New Experimental Attempts at Presenting Mechanization in Problem Solving. *Journal of General Psychology*, 42, 279-297.

Luo, G. (2005). The Relationship Between the Rating Scale and Partial Credit Models and the Implication of Disordered Thresholds of the Rasch Models for Polytomous Response. *Journal of Applied Measurement*, 6/4, 443-455.

Masters, G. N. (1982). A Rasch Model for Partial Credit Scoring. *Psychometrika*, 47/2, 149-174.

Masters, G. N. (1988). Measurement Models for Ordered Response Categories. Der. R. Langeheine ve J. Rost, *Latent Trait and Latent Class Model* (pp. 11-29). New York: Plenum Press.

Masters, G. N. ve Wright, B. D. (1984). The Essential Process in a Family of

Measurement Models. *Psychometrika*, 49/4, 529-544.

Masters, G. N. ve Wright, B. D. (1997). The Partial Credit Model. Der. W. J. van den Linden ve R. K. Hambleton, *Handbook of Modern Item Response Theory* (pp. 101-121). New York: Springer-Verlag New York Inc.

Maxwell, S. E. ve Delaney, H. D. (2004). *Designing Experiments and Analyzing Data: A Model Comparison Perspective* (2. bs.). New York: Psychology Press.

Mcdonald, R. P. (1999). *Test Theory: A Unified Treatment.* New Jersey: Lawrence Erlbaum Association, Inc.

Mellenberg, G. J. (1994). A Unidimensional Latent Trait Model for Continuous Item Responses. *Multivariate Behavioral Research*, 29, 223-236.

Miller, G. J. ve Whicker, M. L. (Der.) (1999). *Handbook of Research Methods in Public Administration.* Marcel Dekker, Inc.

Mooney, C. Z. (1997). *Monte Carlo Simulation* (Sage University Paper Series on Quantitative Applications in the Social Sciences). Thousand Oaks, CA: Sage Publications, Inc.

Muraki, E. (1992). A Generalized Partial Credit Model: Application of an EM Algorithm. *Applied Psychological Measurement*, 16, 159-176.

Muraki, E. (1997). A Generalized Partial Credit Model. Der. W. J. van den Linden ve R. K. Hambleton, *Handbook of Modern Item Response Theory* (pp. 153-164). New York: Springer-Verlag New York Inc.

Myers, J. L. ve Well, A. D. (2003). *Research Design and Statistical Analysis* (2. bs.). New Jersey: Lawrence Erlbaum Associates, Inc.

Olivares, A. M., Drasgow, F. ve Mead, A. D. (1994). Distinguishing Among Parametric Item Response Models for Polychotomous Ordered Data. *Applied Psychological Measurement*, 18/3, 245-256.

Orlando, M. "Critical Issues to Address When Applying Item Response Theory (IRT) Models", (çevrimiçi) http://outcomes.cancer.gov/conference/irt/orlando.pdf, 15 Eylül 2009.

Ostini, R. Nering, M. L. (2006). *Polytomous Item Response Theory Models.* (Sage University Paper Series on Quantitative Applications in the Social Sciences). Beverly Hills: Sage Publications, Inc.

Page, M. C., Braver, S. L. ve MacKinnon, D. P. (2003). *Levine's Guide to SPSS for Analysis of Variance* (2.bs.). New Jersey: Lawrence Erlbaum Associates, Inc.

Punch, K. F. (2005). *Introduction to Social Research–Quantitative and Qualitative Approaches.* London: Sage Publications, Inc.

Quinn, H. (2009). *What is Science?* Physics Today.

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainments Tests.* Chicago: University of Chicago Press.

Reise, S. P. ve Yu, J. (1990). Parameter Recovery in the Graded Response Model Using MULTILOG. *Journal of Educational Measurement*, 27/2, 133-144.

Reise,S. P. ve Waller, N. G. (2002). Item Response Theory for Dichotomous Assessment Data (Der.) F.Drasgow, ve N. Schmitt, *Measuring and Analyzing Behavior in Organizations: Advances in Measurement and Data Analysis* (pp. 88-122). California: Jossey Bass.

Samejima, F. (1969). Estimation of Latent Ability Using a Response Pattern of Graded Scores. *Psychometrika Monograph*, No. 17.

Samejima, F. (1997). Graded Response Model. Der. W. J. van den Linden ve R. K. Hambleton, *Handbook of Modern Item Response Theory* (pp. 85-100). New York: Springer-Verlag New York Inc.

Şencan, H. (2002). *Bilimsel Yazım.* İstanbul: İ.Ü. İşletme Fakültesi.

Şencan, H. (2005). *Sosyal ve Davranışsal Ölçümlerde Güvenilirlik ve Geçerlilik.* Ankara: Seçkin Yayıncılık.

Şenesen, Ü. (2004). *İstatistik: Sayıların Arkasını Anlamak* (2. bs.). İstanbul: Literatür Yayıncılık.

Sharma, S. (1996). *Applied Multivariate Techniques.* Toronto: John Wiley and Sons, Inc.

Shohov, S. P. (2003). *Advances in Psychology Research* (v.7). Nova Publishers.

Si, C. F. B. (2002). *Ability Estimation Under Different Item Parameterization and Scoring Models.* Yayınlanmamış Doktora Tezi. University of North Texas.

SPSS Inc., (2007) *Statistics 17.0 Command Syntax Reference*, Chicago.

Steiger, J. H. ve Fouladi, R. T. (1997). *Noncentrality Interval Estimation and the Evaluation of Statistical Models.* New Jersey: Lawrence Erlbaum.

Stevens, J. P. (2009). *Applied Multivariate Statistics for the Social Sciences* (5. bs.). New York: Routledge.

Stone, M. H. (2004). Substantive scale construction. Der. E. V. Smith Jr. ve R. M. Smith, *Introduction to Rasch measurement* (pp.201–225). Maple Grove, MN: JAM Press.

Thissen, D. (2003). *MULTILOG 7: Multiple categorical item analysis and test scoring using item response theory (Bilgisayar Programı).* Chicago, IL: Scientific Software.

Thissen, D. ve Steinberg, L. (1986). A Taxonomy of Item Response Models. *Psychometrika*, 51, 567-577.

Thompson, B. (2002). What Future Quantitative Social Science Research Could Look Like: Confidence Intervals for Effect Sizes? *Educational Researcher*, 31, 25-32.

Thompson, B. (2006). *Foundation of Behavioral Statistics: An Insight-Based Approach.* New York: The Guilford Press.

Trochim, W. M. (2006). The Research Methods Knowledge Base (2.bs.), (çevrimiçi) http://www.socialresearchmethods.net/kb/, 15 Eylül 2009.

Uebersax, J. S. "Likert scales: dispelling the confusion", Statistical Methods for Rater Agreement website. 2006, (çevrimiçi) at: http://john-

uebersax.com/stat/likert.htm. 10 Şubat 2009.

Weinfurt, K. P. (1995). Multivariate Analysis of Variance. Der. L. G. Grimm ve P. R. Yarnold, *Reading and Understanding More Multivariate Statistics* (pp. 245-276). Washington DC: American Psychological Association.

Weinfurt, K. P. (2000). Repeated Measures Analyses: ANOVA, MANOVA and HLM. Der. L. G. Grimm ve P.R. Yarnold, *Reading and Understanding More Multivariate Statistics* (pp. 317-361). Washington DC: American Psychological Association.

Wilson, M. (2003). On Choosing a Model for Measuring. *Methods of Psychological Research –Online*, 8(3), 1-22.

Wright B. D. (1999). Fundamental Measurement for Psychology. Der. S. E. Embretson ve S. L. Hershberger, *The New Rules of Measurement* (pp. 65-104). New Jersey: Lawrence Erlbaum Associates, Inc.

Wright, B. D. (1992). IRT in the 1990s: Which Model Work Best? *Rasch Measurement Transaction*, 6/1, 196-200.

Wright, B. D. ve Masters, G. N. (1982). *Rating Scale Analysis.* Chicago: Mesa Press.

Wright, B. D. ve Stone, M. H. (1979). *Best Test Design.* Chicago: Mesa Press.

Yıldırım, C. (1979). *Bilim Felsefesi* (11. bs.). İstanbul: Remzi Kitabevi.

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Doğum Tarihi: 1978

Medeni Hali: Bekâr

Ev Adresi: Mehter Çeşme Mh. Cumhuriyet Caddesi 1989 Sk. Kuas Kent C Blok Kat:7

Daire: 29 Beylikdüzü / İstanbul

ÖĞRENİM DURUMU

- İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü, Örgütsel Davranış Doktora Programı, İstanbul, 2009
- İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü, Davranış Bilimleri Yüksek Lisans Programı, İstanbul, 2003
- İstanbul Üniversitesi, İşletme Fakültesi, Lisans, İstanbul, 1999
- İSTEK Vakfı Bilge Kağan Lisesi, Lise, İstanbul, 1995

ASKERLİK DURUMU

Kısa dönem er olarak 2000 yılında tamamlanmıştır.

İŞ DENEYİMİ

- İstanbul Üniversitesi, 2009-Devam
  İşletme Fakültesi, Örgütsel Davranış Ana Bilim Dalı, Öğretim Görevlisi
- İstanbul Üniversitesi, 2001-2009
  İşletme Fakültesi, Örgütsel Davranış Ana Bilim Dalı, Araştırma Görevlisi
- T.G. Ticaret Lisesi, 2000-2001
  Öğretmen, Ekonomi ve İnsan İlişkileri Dersleri
- Zeki Bilardo A.Ş., 1997-2000
  Dış İlişkiler Sorumlusu

AKADEMİK EĞİTİMLER

- Rasch Modelleri (2 Hafta), 2008, University of Western Australia
- Yapısal Eşitlik Modellemesi Eğitimi (3 Gün), 2004, Chicago University

AKADEMİK İLGİ ALANLARI

- Klasik Test Kuramı
- Madde Yanıt Kuramı ve Rasch Analizi
- Çok Değişkenli İstatistik
- Araştırma Tasarımı
- Psikometrik Yöntemler

YABANCI DİL

- İngilizce, (İleri Düzeyde)