

İki Parametrelili Madde Tepki Kuramı Modelinde Parametre Değişmezliğinin İncelenmesi

Burcu HASANÇEBİ^{1*}, Yüksel TERZİ², Zafer KÜÇÜK³

^{1,3}Karadeniz Teknik Üniversitesi, Fen Fakültesi, İstatistik ve Bilgisayar Bilimleri Bölümü, 61080, Trabzon, Türkiye

²Ondokuz Mayıs Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, 55270, Samsun, Türkiye

(Alınış / Received: 27.02.2020, Kabul / Accepted: 04.06.2020, Online Yayınlanma / Published Online: 20.08.2020)

Anahtar Kelimeler

Ölçme,
Madde tepki kuramı,
İki parametrelili model,
Kruskal-Wallis Testi

Özet: Ölçme teorisi incelendiğinde, temel olarak iki kurama rastlamak mümkündür. Bunlardan ilki Klasik Test Kuramı, diğeri ise Madde Tepki Kuramıdır. Madde Tepki Kuramını Klasik Test Kuramından ayıran birçok özellik mevcuttur. Bunlardan biri de parametre değişmezliği özelliğidir. Literatür incelendiğinde değişmezlik özelliğinin hala tartışma konusu olduğu görülmektedir. Bu bilgilerden hareketle bu çalışmada, Karadeniz Teknik Üniversitesi İstatistik ve Bilgisayar Bilimleri Bölümü Nesneye Yönelik Programlama dersinin final sınavı verileri incelenmiştir. Sınav 23 maddeden oluşmaktadır ve dersi alan tüm öğrencilere uygulanmıştır. Analizlere başlamadan önce, veri tarafından en çok desteklenen modelin İki Parametrelili Madde Tepki Kuramı Modeli olduğu belirlenmiştir. Madde parametrelerinin değişmezlik özelliğini incelerken, katılımcılar cinsiyet ve dersi alış sayılarına göre alt bölümlere ayrılmıştır. Kişi parametreleri için ise maddeler, güçlük ve ayırt edicilik indekslerine göre ayrılmıştır. Her bir alt test için gerçek veriye benzer karakterde tepki vektörü simülasyon ile elde edilmiştir. Böylelikle, mevcut olan az miktardaki gerçek veri artırılarak daha doğru sonuçlar elde etmek amaçlanmıştır. Analizden elde edilen sonuçlara göre madde ve kişi parametrelerinin tahminlerinin parametre değişmezliği ilkesine uygun olup olmadığı yorumlanmıştır.

Investigation of Parameter Invariance in Two-Parameter Item Response Theory Model

Keywords

Measurement,
Item response theory,
Two-Parameters model,
Kruskal-Wallis Test

Abstract: When the measurement theory is examined, it is possible to come across two theories. The first of these is Classical Test Theory and the other is Item Response Theory. There are many features that distinguish Item Response Theory from Classical Test Theory. One of these is the parameter invariance feature. When the literature is examined, it is seen that the invariance feature is still the subject of debate. Based on this information, in this study, the final exam data of Karadeniz Technical University Statistics and Computer Science Department Object Oriented Programming course were examined. The exam consists of 23 items and was applied to all students who took the course. Before starting the analyzes, it was determined that the model supported by the data was the Two Parameter Item Response Theory Model. When examining the invariance property of the item parameters, the participants were subdivided according to gender and the number of times they took the course. For the person parameters, the items were divided according to difficulty and discrimination indexes. The response vector for each subtest was obtained by simulation, with a character similar to the real data. According to the results obtained, it is interpreted whether the parameter invariance feature is provided or not.

*İlgili yazar: burcuhsncb@gmail.com

1. Giriş

Ölçme teorisinin tarihine bakıldığında, ilk olarak Klasik Test Kuramı (KTK)'nin etkilerine rastlanmaktadır. KTK'da madde istatistikleri, ölçeğin uygulandığı gruba göre farklılık göstermektedir. Aynı maddelerin ilgili istatistikleri farklı özellikteki gruplara uygulandığında farklı sonuçlar vermektedir. Bu durumun tam aksine, Madde Tepki Kuramında (MTK), elde edilen madde istatistikleri grupların özelliklerinden bağımsızdır. İndeksler, ölçek hangi gruba uygulanırsa uygulansın değişmemektedir. Benzer şekilde, ölçeği alan kişilerin yetenek parametreleri de madde parametrelerinden bağımsız olarak elde edilir [1,2,3]. Yetenek ve madde parametreleri için söz konusu değişmezlik ilkesi ancak MTK varsayımlarının sağlanması ile mümkündür.

MTK'nın kullanılabilmesi için sağlanması gereken varsayımlar; tek boyutluluk, yerel bağımsızlık ve normallik varsayımlarıdır. Tek boyutluluk ölçeğin tek bir örtük özelliği ölçtüğü anlamına gelir. Tek boyutluluğun sağlanması için ölçekte baskın tek bir faktör yapısının gözlemlenmesi beklenir. Yerel bağımsızlık varsayımı, maddelere verilen tepkilerin istatistiksel olarak birbirinden bağımsız olması anlamına gelir. Yapılan tek boyutluluk testinin sonuçlarına göre, tek boyutluluk kriteri 0.9964 bulunmuştur. Yapının tek boyutlu olduğunu söyleyebilmek için tek boyutluluk kriterinin 1'e yakın çıkması gerekmektedir [4]. Tek boyutluluk varsayımının sağlanması halinde yerel bağımsızlık varsayımının da sağlandığı ortaya konulmuştur [3, 5-7]. Son varsayım normallik varsayımıdır. Normallik, KTK ile ortak tek varsayım olup ölçülen özelliğin normal dağılmasına dayanır. Ancak, MTK ile ölçme yapılırken artık lojistik regresyon modelleri kullanıldığından bu varsayım ortadan kalkmıştır.

Değişmezlik özelliği günümüzde tartışma konusu olmaya devam etmektedir. Bazı çalışmalarda değişmezlik sağlanırken bazısında belirli parametreler için sağlanmıştır. Bazı çalışmalarda ise bu ilke sağlanamamıştır. Doğan ve Kılıç, 3 parametrelili model (3PL) ile yapılan çalışmalarında yetenek parametresinin değişmezlik ilkesini sağladığını ortaya koymuşlardır. Bunun yanında çalışmada madde parametreleri için yapılan analiz sonucunda, ayırıcılık ve şans parametresine göre güçlük parametresinin değişmezliği daha iyi sağladığı sonucuna varılmıştır [8]. Sünbül ve Erkuş, değişmezlik ilkesinin farklı örnek büyüklüklerinden ne derecede etkilendiğini araştırmış, *a* ve *b* parametrelerinin değişmezlik özelliğinin büyük ölçüde sağladığı sonucuna ulaşmıştır [9]. Reyhanlıoğlu Keçeoğlu, farklı boyut ve örnek büyüklüklerinde değişmezlik ilkesi incelenmiş, yapının tek boyutlu ya da çok boyutlu olmasıyla örnek büyüklüklerinin farklı olması durumlarında değişmezlik ilkesinin etkilendiğini ortaya koymuştur

[10]. Acar ve Kelecioğlu, Genelleştirilmiş Aşamalı Doğrusal Model (GADM) ve Rasch modelden tahmin edilmiş kişi ve madde parametrelerinin değişmezlik özelliğini incelemişler ve iki modelde de madde parametreleri yüksek düzeyde değişmez iken, yetenek parametresi için değişmezliğin GADM ile tam olarak sağlandığını öne sürmüşlerdir [11]. Kelkar vd., Tıp Fakültesi sınavları için sadece güçlük parametresi içeren 1 parametrelili (1PL) model, güçlük ve ayırt edicilik parametrelerini içeren 2 parametrelili (2PL) model ve güçlük ayırt edicilik ile şans parametresi içeren 3 parametrelili (3PL) MTK modellerinde değişmezlik ilkesini incelemişlerdir. Madde parametreleri için cinsiyet, etnik grup ve dile göre alt gruplar oluşturmuş, yetenek parametresinin değişmezliği için ise güçlğe göre alt testlere ayrılmıştır. 1PL'den elde edilen tahminler en küçük standart hataya sahipken, 2PL ve 3PL modeller arasındaki fark oldukça küçük bulunmuştur. Tüm modeller için değişmezlik sağlanmasına rağmen en güçlü değişmezlik 1PL modelde sağlanmıştır [12]. Custer vd., 1PL ve 3PL modelde değişmezlik ilkesini incelemişlerdir. Veriden en fazla desteği alan modelin 3PL olmasına rağmen değişmezlik ilkesini düşük, orta ve yüksek yetenek gruplarının hepsinde ve madde parametrelerinde 1PL model sağlamıştır [13]. Adedoyin, yetenek parametresinin değişmezliğini KTK ve MTK anlamında karşılaştırmıştır. KTK ile tahmin edilen kişi parametrelerinde değişmezlik gözlenmezken, MTK ile yapılan tahminlerde değişmezlik özelliği saptanmıştır [14]. Adedoyin vd., madde parametrelerinin değişmezliği MTK ve KTK'ya dayalı araştırmışlardır. KTK ile yapılan analizler sonucu güçlük parametresinin örnekten örneğe farklılık gösterdiğini, MTK ile yapılan tahminlerin ise farklı örneklerde değişmez olduğunu ortaya koymuşlardır [15].

Bu çalışmada, Nesneye Yönelik Programlama dersinin final sınavı verileri üzerinde değişmezlik çalışması yapılmıştır. Analizler hem madde hem kişi parametreleri için yapılmış, her durumda veri çeşitli özelliklere göre alt testlere bölünmüştür. Ölçek, 81 kişiye uygulanmıştır. Ancak, elde edilen alt testlerde çok az sayıda örnek bulunması sebebiyle her alt test için, gerçek veriye benzer karakterde cevap verisi simülasyon ile elde edilmiştir. Böylelikle, her bir alt test için 1000'er kişilik tepki vektörü oluşturulmuştur. Literatürde, model veri uyumunun en yüksek olduğu modeller genelde 3PL model olarak seçilmiş ancak, değişmezliğin en iyi 1PL modelle sağlandığı görülmüştür. Çalışmada, veriden en yüksek desteği alan model 2PL model olarak belirlenmiş ve 2PL modelde değişmezlik ilkesi incelenmiştir.

2. Materyal ve Metot

Bu çalışmada, Karadeniz Teknik Üniversitesi İstatistik ve Bilgisayar Bilimleri Bölümü 2018-2019 Eğitim Öğretim yılı Bahar döneminde verilen Nesneye

Yönelik Programlama dersinin final sınavı analiz edilmiştir. Ölçek, 23 maddeden oluşmaktadır. Her bir madde 4 seçenektir. Test, dersin kazanımları baz alınarak hazırlanmıştır. Ölçek, dersi alan 81 öğrenciye uygulanmış ve öğrencilerden elde edilen veriler 0-1 şeklinde ikili sisteme dönüştürülmüştür. Yetenek ve madde parametrelerinin her biri için deęişmezlik ilkesi incelenmiştir. Bunu yaparken, herhangi bir cinsiyet grubunun veya dersi bir ya da birden fazla kez almış olmanın parametre tahmininde etkili olup olmadığını araştırabilmek için madde parametrelerinin deęişmezlięi incelenirken öğrenciler, cinsiyetleri ve dersi alış sayılarına göre gruplandırılmıştır. Benzer şekilde, yetenek parametresinin deęişmezlięi için ise ölçek, maddelerin güçlük ve ayırt edicilik düzeylerine göre alt testlere bölünmüştür. Böylelikle, benzer ya da farklı maddelerin yetenek parametresinin tahminini deęiştirip deęiştirmeyeceęi incelenmiştir. Mevcut tüm alt testler için gerçek veriye benzer karakterde yapılan simülasyon çalışması ile tüm alt testlerde örnek sayısı 1000'e çıkarılmıştır.

Tablo 1. Grup ve test için betimsel istatistikler

Örnek Sayısı	Madde Sayısı	Dizi Genişlięi	Ortalama Güçlük
8000	23	22	0.05

Tablo 1'de grup ve test için bazı betimsel istatistikler verilmiştir. Ölçek 81 kişiye uygulanmış ancak yapılan simülasyon çalışması ile örnek sayısı 8000'e çıkarılmıştır. Testin ortalama güçlüğü 0.05 olarak hesaplanmıştır. Testin ortalama güçlüğü 0.5'den küçük olması, ölçekteki maddelerin öğrenciye zor geldięi ya da düşük düzeyli öğrencilerin çoğunlukta olduęu biçiminde yorumlanabilir. Ölçek 23 maddeden oluşmaktadır ve ranj aralıęı 22 olarak hesaplanmıştır.

Ölçme teorisinde kullanılan bir ölçeğin geçerlilięinin ve güvenilirlięinin düşük olması da analizlerden elde edilecek kestirimlerin objektiflięine, hassaslıęına ve doğruluęuna şüphe düşürmektedir. Buradan doğan ihtiyaçla, mevcut veriye güvenilirlik analizi yapılmış ve Cronbach alfa katsayıları her bir madde için 0.92 civarında bulunmuştur. Böylelikle, geliştirilen ölçeğin yüksek düzeyde güvenilir olduęunu söylemek mümkündür.

Analizlerden önce modele karar vermek en önemli adımdır. Verilerden en yüksek desteęi alan modeli seçmek yapılacak kestirimlerin doğruluęu açısından oldukça önemli bir rol oynar. Bu sebeple, 1PL, 2PL ve 3PL modellerin çeşitli uyum kriterleri ile karşılaştırmaları yapılmıştır. Elde edilen sonuçlara göre, AIC, BIC ve DIC deęerlerinin en küçük, log likelihood deęerinin ise en büyük olduęu model olan 2PL modelin en uygun model olduęu Tablo 2'de verilmiştir. 2PL modelde madde parametresi olarak

güçlük parametresinin yanında ayırt edicilik parametresi de mevcuttur. Kişi parametresi olarak yetenek parametresi her modelde bulunmaktadır.

$$P_i(\theta) = \frac{\exp[a_i(\theta - b_i)]}{1 + \exp[a_i(\theta - b_i)]} \quad (1)$$

Eşitlik 1. ile 2PL model matematiksel olarak ifade edilmektedir [16]. Burada, i , madde sayısı olmak üzere; a_i , i . maddenin ayırt edicilik parametresidir. Ayırt edicilik parametresi, maddenin ölçülmek istenen özellięe sahip olan öğrenci ile olmayan öğrenciyi birbirinden ayırabilme becerisidir. b_i , i . maddenin güçlük indeksidir. Bu indeks ile, maddenin zorluk derecesi hakkında kolay, orta ya da zor biçiminde yorum yapılmasını sağlar. Kişi parametresi θ ise öğrencilerin yetenek düzeylerini gösteren parametredir.

Kişilere uygulanan ölçeklerin ana amacı ilgili kişilerin ölçülmek istenen örtük özelliklere sahip olup olmadıklarını, sahipler ne kadar sahip olduklarını belirlemektir. Böylece, kişilerin yetenek ölçeęi üzerindeki konumları belirlenir. Yetenek parametrelerinin tahmini için madde parametrelerinin bilinmesi gerekmektedir. Kişilerden elde edilen cevap vektörü ve madde parametreleri yardımıyla kişi yeteneklerinin tahminleri yapılır [8]. Yetenek parametresinin kestirimi yapılırken öncelikle ham puanlar mı tüm cevapların örüntüsü mü kullanılacak buna karar verilmelidir. Örüntü yönteminde, aynı sayıda fakat farklı güçlük düzeylerindeki maddeleri doğru cevaplayanlar farklı yetenek tahminine sahip olabilmektedirler. Toplam puan ile toplam doğru cevap sayıları kullanılarak yetenek kestiriminde bulunmaktadır. Bu yöntemde aynı doğru cevap puanına sahip öğrenciler hangi maddeleri doğru yanıtlarsa yanıtlasınlar her zaman aynı yetenek düzeyine sahip olmaktadırlar. Bu nedenle, toplam puan, örüntü puanlamaya göre daha fazla ölçme hatası içerir [17].

Analizlere başlamadan önce tüm tepki vektöründen madde ve kişi parametreleri kestirilmiştir. Parametre deęişmezlięi özellięini incelemek için farklı özelliklere sahip madde ve kişileri bir araya toplayarak alt testler ve alt gruplar oluşturulmuştur. Tüm bu ayrıştırma önceden kestirilmiş parametrelere dayanarak yapılmıştır. Model 2PL olarak seçildięinden modelde güçlük ve ayırtıcılık olarak 2 madde parametresi mevcuttur. Madde parametreleri için deęişmezlik incelenirken kişilerin cinsiyetleri ve dersi kaçınıcı alışları (ilk defa alanlar ve alttan alanlar) olduęuna göre alt gruplara ayrıştırma yapılmıştır. Kişi parametresi (yetenek parametresi) için de güçlükleri

Tablo 2. Model veri uyumu

Model	logLikelihood	AIC	BIC	DIC
1PL/Rasch	-907.5408	1863.082	1920.548	1815.082
2PL	-831.5121	1757.024	1869.563	1663.024
3PL	-853.9027	1799.805	1909.950	1707.805

düşük ve yüksek ile ayrıcalıkları düşük ve yüksek olmak üzere ölçek alt testlere bölünmüştür.

Parametrelerin özelliklerine göre tüm alt gruplar ve alt testlerin oluşturulma adımının ardından elde edilen farklı özellikteki grupların değişmezlik özelliğini sağlayıp sağlamadığı incelenmiştir. Örneğin, güçlük parametresinin değişmezliği incelenirken mevcut maddelerin parametreleri bölünmüş her alt grupta aynı mıdır? Yoksa örnekten örneğe istatistiksel olarak anlamlı farklılıklar göstermekte midir? Bu soruların yanıtı için kestirilen parametrelerin gruptan gruba değişip değişmediği incelenmelidir.

Mevcut veriden elde edilen her bir alt test ve alt gruba normallik testi yapılmıştır. Verinin normal dağılıp dağılmadığının incelenmesi için Shapiro-Wilk Testi kullanılmıştır. Shapiro-Wilk Testinin sonuçlarına göre güçlük, ayırt edicilik ve yetenek parametrelerinin değişmezliğinin analizi için elde edilmiş alt gruplardan kestirilen parametrelerin normal dağılmadığı sonucuna ulaşılmıştır. Bu durumda, alt gruplar arasında fark var mı sorusunun cevabını aramak için parametrik olmayan testlerden Kruskal-Wallis Testi uygulanmış, elde edilen sonuçlara göre değişmezlik özelliği hakkında yorumlar yapılmıştır.

3. Bulgular

Bu çalışmadaki ana amaç, her bir alt grup için kestirilen parametrelerin gruptan gruba değişip değişmediğini ölçmektir. Bu amaca yönelik yapılan ilk çalışma belirlenmiş grupların normal dağılıp dağılmadığını ölçmektir. Bunun için Shapiro-Wilk Testi her bir gruba uygulanmıştır.

Analizlere başlamadan önce uygulanan ölçek yapısının tek boyutlu olduğu tek boyutluluk kriterinin 0,9964 bulunmasıyla belirlenmiş ve modelin Tablo 2'deki uyum kriterlerinden hareketle 2PL modele uyumlu olduğu tespit edilmiştir. Analizler yapılırken bu bilgiler baz alınarak yapılmıştır. Tüm analizler R yazılımı ile gerçekleştirilmiştir.

Tablo 3. Güçlük parametresi için Shapiro-Wilk Test sonuçları

Alt gruplar	Test İstatistiği	p
Kız	0.7704	0.0001
Erkek	0.9029	0.0291
İlk defa alanlar	0.9529	0.3357
Altan alanlar	0.7574	8.591e-05

Tablo 3'de ilgili örneklerin normal dağılımdan geldiği yokluk hipotezini test eden ve anlamlılık düzeyi 0.05 olarak seçilmiş Shapiro-Wilk testinin sonuçları

görülmektedir. Bu durumda, güçlük parametresi için elde edilen sonuçlardan sadece ilk defa alanlar grubunun normal dağılıma sahip olduğu görülmektedir. Geride kalan kız, erkek ve altan alanlar grupları normal dağılımdan gelmemektedir. İlgili örneklerin normal dağılımdan farkı yoktur, yokluk hipotezinin testinde anlamlılık düzeyi 0.05 olarak belirlenmiştir ve Tablo 4'de verilen p değerlerine bakıldığında, örnek içerisindeki tüm grupların normal dağılımdan gelmediğini görmek mümkündür.

Tablo 4. Ayrıcalık parametresi için Shapiro-Wilk Test sonuçları

Alt gruplar	Test İstatistiği	p
Kız	0.5607	3.441e-07
Erkek	0.4220	1.671e-08
İlk defa alanlar	0.49617	7.881e-08
Altan alanlar	0.63435	2.199e-06

Tablo 5. Yetenek parametresi için Shapiro-Wilk Test sonuçları

Alt gruplar	Test İstatistiği	p
Güçlükler düşük	0.9174	6.24e-23
Güçlükler yüksek	0.9002	5.46e-25
Ayrıcalık düşük	0.9031	1.16e-24
Ayrıcalık yüksek	0.8330	4.51e-31

Tablo 5'de örneklerin normal dağılımdan gelip gelmediğini test eden Shapiro-Wilk testinin test istatistikleri ve p değerleri görülmektedir. Elde edilen sonuçlara göre hiçbir grup normal dağılımdan gelmemektedir.

Yapılan normallik testlerinden elde edilen bilgilere göre çoklu gruplar arasında istatistiksel olarak anlamlı bir fark bulunup bulunmadığının testi için parametrik olmayan testlere başvurulmalıdır. Parametrik testlerden ANOVA'nın karşılığı olan Kruskal-Wallis testi bu veriler için uygun test olacaktır.

Tablo 6. Güçlük parametresi için Kruskal-Wallis Testi sonuçları

Kruskal-Wallis rank sum test		
χ^2	Serbestlik derecesi	p
2.8368	3	0.4175

Tablo 6 ile güçlük parametresinin değişmezliği incelenmiştir. Değişmezlik incelenirken gruplar arasında fark olup olmadığına Kruskal-Wallis testi ile bakılmıştır. Test, kestirilen parametrelerin gruptan gruba değişmediğini ifade eden yokluk hipotezini test eder ve anlamlılık düzeyi 0.05 olarak belirlenmiştir. Tablo 6'da gösterilen p değerine göre yokluk hipotezi kabul edilir. Bu gruplar arasında kestirilen güçlük parametrelerinin arasında istatistiksel olarak anlamlı

bir fark yok olarak yorumlanır. Yani, güçlük parametresi için değişmezlik ilkesi sağlanmıştır.

Tablo 7. Ayırıcılık parametresi için Kruskal-Wallis Testi sonuçları

Kruskal-Wallis rank sum test		
χ^2	Serbestlik derecesi	p
51.517	3	3.796e-11

Tablo 7'de ayırt edicilik parametresinin belirlenmiş alt gruplar için değişmezlik ilkesini sağlayıp sağlamadığına ilişkin test sonuçları görülmektedir. Anlamlılık düzeyi 0.05 olmak üzere; p değerine bakıldığında ayırıcılık indeksinin değişmezlik özelliğini sağlamadığı görülmektedir. Bu durumda; hangi alt grupta farklılık olduğunun belirlenmesi için Kruskal-Wallis testi için çoklu karşılaştırmalar Dunn Testi ile yapılmıştır.

Tablo 8. Ayırıcılık parametresi için Dunn Testi sonuçları

Satır ortalaması / Sütun ortalaması		Alttan alanlar	Erkek	Kız
Erkek	Test ist. p	-5.6432 0.0000*		
Kız	Test ist. P	-5.7150 0.0000*	-0.0717 1.0000	
İlk defa alanlar	Test ist. p	-6.1677 0.0000*	-0.5245 1.0000	-0.4527 1.0000

Tablo 8'de ayırıcılık parametresi için Kruskal-Wallis'in çoklu karşılaştırma testi yapılmıştır. Elde edilen sonuçlara göre, erkek ve alttan alanlar grubu, kız ve alttan alanlar grubu ve ilk defa alanlar ve alttan alanlar grubu arasında istatistiksel olarak anlamlı fark bulunmuştur. Bu durumda, bu alt gruplar için kestirilen ayırıcılık parametresinin değişmezlik ilkesini sağlamadığı söylenebilir. Yani, bu gruplarda örnekten örneğe kestirilen ayırıcılık değerleri farklı olacaktır.

Tablo 9. Yetenek parametresi için Kruskal-Wallis Testi sonuçları

Kruskal-Wallis rank sum test		
χ^2	Serbestlik derecesi	p
242.29	3	2.2e-16

Tablo 9'da yetenek parametresi için yapılmış Kruskal-Wallis testinin sonuçları görülmektedir. Belirlenmiş alt gruplar arasında istatistiksel olarak anlamlı bir fark yoktur yokluk hipotezini test eden testin p değerine göre en az bir grupta farklılık olduğu sonucuna varılmıştır. Hangi gruplar arasında farklılık olduğunu belirlemek için çoklu karşılaştırma testlerinden Dunn Testi kullanılmıştır.

Tablo 10. Yetenek parametresi için Dunn Testi sonuçları

Satır Ortalaması / Sütun Ortalaması		Ayırıcılığı yüksek	Güçlüğü düşük	Güçlüğü yüksek
Güçlüğü düşük	Test ist. p	-11.3187 0.0000*		
Güçlüğü yüksek	Test ist. P	-10.7772 0.0000*	0.5414 1.0000	
Ayırıcılığı düşük	Test ist. p	-14.6340 0.0000*	-3.3153 0.0027*	-3.8568 0.0000*

Kruskal-Wallis Testinde bazı gruplarda istatistiksel olarak anlamlı farklılık gözlemlendiği Tablo 9 ile belirlenmiştir. Tablo 10'da ise bu anlamlı farklılığın hangi gruplar arasında olduğunu belirlemek için Dunn Testi yapılmış ve sonuçları gösterilmiştir. Tablo 10'a göre, ayırıcılığı yüksek ile güçlüğü düşük grup, güçlüğü yüksek grup ve ayırıcılığı yüksek grup ayrıca da güçlüğü düşük grupla ayırıcılığı düşük grup arasında yapılmış kestirimlerde anlamlı farklılıklar gözlemlenmiştir. Bu durumda, belirlenmiş bu gruplardan elde edilen yetenek kestirimleri değişmezlik ilkesini sağlamayacaktır sonucuna ulaşılır.

4. Tartışma ve Sonuç

Bu çalışmada, Karadeniz Teknik Üniversitesi İstatistik ve Bilgisayar Bilimleri Bölümünde verilen Nesneye Yönelik Programlama dersinin final sınavı verileri kullanılmıştır. Değişmezlik ilkesinin incelenmesi için yetenek ve madde parametrelerine göre alt gruplar oluşturulmalı ve bu alt gruplardan elde edilen parametre kestirimlerinin değişmez olup olmadığı incelenmelidir. Bu sebeple, madde parametrelerinin değişmezliğinin incelenmesi için cinsiyet ile dersi ilk defa alanlar ve alttan alanlar şeklinde alt gruplar oluşturulmuştur. Yetenek parametresinin değişmezliğinin incelenmesi için ise güçlüğü düşük, güçlüğü yüksek ve ayırıcılığı düşük ve ayırıcılığı yüksek olmak üzere alt testler oluşturulmuştur.

Nesneye Yönelik Programlama dersini alan toplamda 81 öğrenci mevcuttur ve alt gruplar oluşturulduğunda bu sayı daha da azalmaktadır. Bu sebeple, alt gruplara ve alt testlere ayrılmış veri üzerinden, gerçek veri karakterine benzer veri, simülasyon ile elde edilmiştir. Böylelikle her bir alt grupta gerçek veriye benzeyen 1000'er örnek elde edilmiştir. Yapılan tüm analizler R yazılımı ile yapılmıştır.

MTK modelleri ile çalışabilmek için birtakım varsayımların sağlanması gerekmektedir. Bunlardan ilki tek boyutluluk varsayımdır. Bu sebeple veriye Tek Boyutluluk Testi uygulanmış tek boyutluluk kriteri 0.9964 olarak elde edilmiştir. Literatüre bakıldığında, tek boyutluluktan söz edebilmek için hesaplanan kriterin 1'e yakın olması gerektiği görülmektedir. Kriter 1'e ne kadar yakınsa yapının tek boyutlu olduğunun ispatı o kadar güçlüdür [4]. Diğer bir varsayım yerel bağımsızlık varsayımdır. Bir yapının tek boyutluluk varsayımını sağladığı durumda, yerel bağımsızlık varsayımını da karşıladığını söylemek mümkündür [3, 5, 7]. Son varsayım normallik, kullanılan modellerin lojistik regresyon modelleri olması sebebiyle önemini yitirmiş bir varsayımdır.

Varsayımların sağlanmasıyla birlikte hangi MTK modeli ile çalışılacağı sorusu gündeme gelmektedir.

Model-veri uyumunun maksimum olduęu modeli bulmak yapılacak analiz sonularının doęruluęu iin olduka nemlidir. Aksi halde, yanlış kestirimler yapılır ve bundan kaynaklanan yanlış yorumlar kaçınılmaz olur. Bu durumun nune gemek aısından alıřmada MTK modelleri eřitli uyum kriterleri aısından karřılařtırılmıř ve en uygun modelin 2PL model olduęu grlmřtr. Ardından, modelde mevcut olan glk, ayırıcılık ve yetenek parametrelerinin kestirimi yapılmıřtır.

Kestirimleri yapılmıř model parametrelerinin normallięi arařtırılmıř, parametrelere Shapiro-Wilks Testi uygulanmıřtır. Test sonularına gre, sadece glk parametresi iin yapılan testte dersi ilk defa alanlar grubu normal daęılım gstermiř, dięer gruplarda normal daęılıma rastlanmamıřtır. Hatta, ayırıcılık iin yapılan testte ilk defa alanlar grubu da normal daęılım gstermemiřtir. Bu durumda, gruplardan kestirilmemiř madde ve kiři parametreleri arasında fark olup olmadıęının test edilebilmesi iin Kruskal-Wallis Testi yapılmıřtır. Kruskal-Wallis testi sonularına gre, sadece glk parametresinin deęiřmezlik ilkesinin saęlandıęı gzlemlenmiřtir. Ayırıcılık ve yetenek parametreleri deęiřmezlik zellięini saęlamamıřtır. Bu durumda, ayırıcılık ve yetenek parametrelerinde istatistiksel olarak anlamlı farklılık ieren grupların hangisi olduęu sorusu akla gelmektedir. Kruskal-Wallis testinin oklu karřılařtırma testi Dunn Testidir. Ayırıcılık ve yetenek parametreleri iin Dunn Testi yapılmıř ve sonular sırasıyla Tablo 8 ve Tablo 10'da verilmiřtir.

Literatr incelendięinde, deęiřmezlik ilkesi ile ilgili farklı sonular alınmıř alıřmalar olduęu grlmektedir. 3PL modellerin veri uyumlarının en yksek olduęu ancak deęiřmezlięin 1PL modeller ile en iyi saęlandıęı ortaya konulmuřtur [13]. Madde parametrelerinden glęu en iyi saęlayan parametrenin ise glk parametresi olduęu belirlenmiřtir. Bunun yanında yetenek parametresinin de deęiřmez olduęu gsterilmiřtir [9]. MTK modelleri ile yapılan deęiřmezlik alıřmasında en gl deęiřmezlięin 1PL model ile saęlandıęı sonucuna varılmıřtır [12]. Glk parametresi iin KTK ve MTK erevesinde deęiřmezlięin incelendięi alıřmada, KTK kapsamında kiři parametresi tahminleri, hibir alt testte deęiřmezlik gstermemiřtir. Yine KTK erevesinde glk parametresinin deęiřmezlięi iin farklı alt gruplarda deęiřmezlięin saęlandıęı sonucuna varılmıřtır. MTK erevesinde yapılan tahminlerde ise, glk parametresi farklı alt gruplarda deęiřmez iken, farklı rnek boyutlarında deęiřmezlięi saęlamadıęı sonucu ortaya konulmuřtur [15]. Yapılan literatr arařtırmaları sonunda konuyla ilgili 2PL modelle yapılmıř alıřmaların sayıca fazla olmadıęı genelde 1PL model ve 3PL modeller ile alıřıldıęı, bazı durumlarda iki modelin karřılařtırmasının yapıldıęı gze arpmaktadır. Ayrıca, bu alıřma gerek veriden hareketle simlasyon verisi retmesi sebebiyle sadece

simlasyon ile elde edilmiř veri ile yapılan alıřmalardan ve sadece gerek veri ile yapılmıř alıřmalardan ayrılmaktadır. Bu zellikleri ile yapılan alıřmada sadece glk parametresinin deęiřmezlięi saęladıęı sonucuna ulařılmıřtır.

Kaynaka

- [1] Hambleton, R., Swaminathan, H. 1985. Item Response Theory: Principles and Applications. Springer Science& Business Media, USA, 332s.
- [2] Embretson, S.E., Reise, S.P. 2000. Item Response Theory for Psychologists. Psychology Press, Los Angeles, 384s.
- [3] De Mars, C. 2010. Item Response Theory. Oxford University Press, New York, 130s.
- [4] Drasgow, F., Lissak, R. I. 1983. Modified parallel analysis: A Procedure for Examining the Latent Dimensionality of Dichotomously Scored Item Responses. Journal of Applied Psychology, 68(3), 363-373.
- [5] De Gruijter, DNM., van der Kamp, LJT. 2008. Statistical Test Theory for the Behavioral Sciences. Psychometrika, 73(4), 791-792.
- [6] Recklase, M.D. 2009. Multidimensional Item Response Theory. Springer, New York, 339s.
- [7] Gler, N. ed. 2016. Madde Tepki Kuramının Temelleri. Pegem Akademi, Ankara, 171s.
- [8] Doęan, N., Kılı, A.F. 2017. Madde Tepki Kuramı Yetenek ve Madde Parametreleri Kestirimlerinin Deęiřmezlięinin İncelenmesi. ss 297-314. Demirel, ., Diner, S., ed. 2017. Kreselleřen Dnyada Eęitim, Pegem Akademi, Ankara, 902s.
- [9] Snbl, ., Erkuř, A. 2013. Madde Parametrelerinin Deęiřmezlięinin eřitli Boyutluluk zellięi Gsteren Yapılarda Madde Tepki Kuramına Gre İncelenmesi. Mersin niversitesi Eęitim Fakltesi Dergisi, 9(2), 378-398.
- [10] Reyhanlıoęlu Keceoęlu, . 2018. Parametrik ve Parametrik Olmayan Madde Tepki Kuramında Farklı rneklem Byklklerine ve Boyutluluklarına Gre Parametre Deęiřmezlięinin İncelenmesi. Hacettepe niversitesi, Eęitim Bilimleri Enstits, Doktora Tezi, 111s, Ankara.
- [11] Acar, T., Kelecioęlu, H. 2008. Genelleřtirilmemiř Ařamalı Doęrusal Model ile Rasch Modelinin Parametre Deęiřmezlięinin Karřılařtırılması. I. Ulusal Eęitimde ve Psikolojide lme ve Deęerlendirme Kongresi, 14-16 Mayıs, Ankara, 181-193.
- [12] Kelkar, V., Wightman, L.F., Luecht, R.M. 2000. Evaluation of the IRT Parameter Invariance Property for the MCAT. Annual Meeting of the

National Council on Measurement in Education,
25-27 April, New Orleans.

Statistical theories of mental test score, Addison-
Wesley, USA, 460s.

- [13] Custer, M., Sharairi, S., Yamazaki, K., Signatur, D., Swift, D., Frey, S. 2008. A Paradox Between IRT Invariance And Model-Data Fit When Utilizing the One-Parameter and Three-Parameter Models. Annual Meeting of the American Educational Research Association, 24-28 March, New York, 70-71.
- [14] Adedoyin, O.O. 2010. Investigating the Invariance of Person Parameter Estimates Based on Classical Test and Item Response Theories. International Journal of Education Science, 2(2), 107-113.
- [15] Adedoyin, O.O., Nenty, H.J., Chilasa, B. 2008. Investigating the Invariance of Item Difficulty Parameter Estimates Based on CTT and IRT. Educational Research and Review, 3(2), 83-93.
- [16] Birnbaum, A. 1968. Some Latent Trait Models and Their Use in Inferring an Examinee's Ability. ss 379-398. Lord, F.M., Novick, M.R. ed. 1968.
- [17] Orlando, M., Thissen, D. 2000. New Item Fit Indices for Dichotomous Item Response Theory Models. Applied Psychological Measurement, 24(1), 50-64.