

Sato Test Kuramı Yöntemleriyle Farklı Örneklemelerden Elde Edilen Madde Parametrelerinin Tutarlılığının İncelenmesi

Examination of the Consistency of the Sato Test Theory Item Parameters Obtained From Different Samples

Sait ÇÜM *

Selahattin GELBAL**

Ching-Pin TSAI***

Öz

Bu araştırmada, aynı evrenden tesadüfi olarak seçilen örneklemelerden elde edilen Maddeler için Sınırlandırılmış Gri İlişki Derecelerinin (MİSGİD) tutarlılığının ve Madde Uyarı İndeksleri (MUİ) aracılığı ile gerçekleştirilen madde sınıflandırmaları arasındaki uyumun incelenmesi amaçlanmıştır. Bu amaç doğrultusunda, 20 madde ve bu maddelere dayalı 1000 adet yanıt örüntüsü simülasyon yoluyla elde edilmiş, yanıt örüntüleri içerisinde seçkisiz olarak seçilen ve 30 yanıt örüntüsü içeren 10 farklı veri seti (örneklem) oluşturulmuştur. Örneklemelerin her birinden elde edilen MİSGİD değerlerinin farklı örneklemeler arası tutarlılığının belirlenebilmesi amacıyla Spearman Sıra Farkları Korelasyon Katsayıları hesaplanmış ve buna ek olarak değer ortalamaları arasındaki farkın anlamlılığı Kruskal Wallis-H Testi ile incelenmiştir. Ayrıca örneklemelerden elde edilen MUİ ile yapılan sınıflandırmaların ne oranda uyum gösterdiği yüzde oranlar yardımıyla incelenmiştir. Elde edilen bulgular, Sato Test Kuramında maddelerin kalitesinin belirlenebilmesi için üretilen MİSGİD parametrelerinin ve MUİ parametresine dayalı yapılan sınıflandırmaların tutarlı sonuçlar verdiğini göstermiştir.

Anahtar Kelimeler: Sato test kuramı, öğrenci-problem çizelgesi, madde analizi.

Abstract

On this research, it has been aimed to the examination of consistency of Localized Grey Relational Grade-Problem (LGRG-P) values and accommodation item classifications which are made via Problem Caution Index (PCI) between different samples that are randomly chosen from the same universe. In accordance with this purpose, 20 items and 1000 response patterns based on these items have been obtained by simulation, 10 different dataset (sample) which are randomly chosen among response patterns and involve 30 response patterns have been constructed. In order to examination of consistency of obtained LGRG-P values among different samples, Spearman Rank-Order Correlation Coefficients have been calculated and additionally significance of difference among value averages has been examined by Kruskal Wallis-H Test. Moreover, with the help of percentages at what rate of item classifications made by PCI obtained from samples show accommodation. Findings which are obtained have showed that Sato Theory Parameters which are produced in order to determine the quality of items have given consistent results.

Keywords: Sato test theory, student-problem chart, item analysis.

GİRİŞ

Modern toplumda eğitimde ve psikolojide kullanılan testler bireylerin hayatında çok önemli bir rol oynamaktadır. Test puanları, bireylerin eğitim görecekları okulları büyük ölçüde belirlemede dolayısıyla uzmanlaşacakları mesleklerin belirlenmesi sürecinde kritik bir unsur olarak ortaya çıkmaktadır. Başka bir bakış açısıyla test puanları, kişinin hayatı boyunca kendisini nasıl göreceği ve başkalarının onu nasıl göreceğini belirleyen bir etmen haline gelmiştir. Bu anlamda test puanlarının nasıl elde edileceği ve nasıl yorumlanacağı konuları üzerinde dikkatle durulmalıdır. Eğitimde ve

*Doktora Öğrencisi, Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara-Türkiye, e-posta:saitcum@hotmail.com

** Prof. Dr., Hacettepe Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Ankara-Türkiye, e-posta:sgelbal@gmail.com

*** Dr., National Taichung University, Eğitimde Ölçme ve İstatistik Enstitüsü, Taichung-Tayvan, e-posta: tsai.chenbin@msa.hinet.net

psikolojide test geliştirme ve ölçme süreçleri, oldukça karmaşık ve geniş kapsamlı bilgiler bütünü üzerine inşa edilmiş bir bilim dalı olan psikometri çerçevesinde ele alınır. Psikometristler, test maddelerinin nasıl geliştirileceği, nasıl seçileceği, verilen cevapların sayılara nasıl dönüştürüleceği, bunların nasıl yorumlanacağı ve elde edilen verilerin kalitesinin nasıl değerlendirileceği gibi sorulara matematiksel ve istatistiksel teorilere dayalı olarak yanıt ararlar. Söz konusu teorilere “test kuramı” ya da “psikometrik kuram” adı verilmektedir (Suen, 1990).

Klasik Test Kuramı (KTK) ve Madde Tepki Kuramı (MTK) geçmişten günümüze üzerinde en fazla çalışılan başlıca iki test kuramı olarak öne çıkmaktadır. Alanyazında, Klasik Test Kuramı ve Madde Tepki Kuramının bazı üstün ve zayıf yönlerinin tartışıldığı görülmektedir. Hambleton, Swaminathan ve Rogers (1991), Klasik Test Kuramının sınırlılıklarını şu şekilde açıklamaktadırlar: Madde ve test istatistikleri gruba bağlı olarak hesaplanır. Bu sınırlılık madde ve test özelliklerinin birbirinden ayıramamasıyla ilişkilidir. Eğer bir test zor ise kişi düşük yeteneğe, eğer kolay ise yüksek yetenek düzeyine sahipmiş gibi görünebilir. Katılımcıların yeteneği ise test maddelerinin güçlük düzeyini belirlemektedir çünkü KTK’da maddelerin ayırt ediciliği ve güçlüğü belli bir grup için geçerlidir. Bu nedenle madde parametreleri uygulamanın yapıldığı gruba bağlı olarak değişecektir. Bu durum paralel formlar oluşturma ve farklı testlere katılan bireyleri karşılaştırma konularında sorunların yaşanmasına neden olur. Ayrıca KTK’ya göre doğru yanıtlanan toplam soru sayısından elde edilen puan bireyin yetenek düzeyi olarak kabul edilir. Bu tür bir hesaplamada farklı sorulara doğru yanıt vermiş olmalarına rağmen eşit sayıda doğru yanıt bulunan bireyler aynı yetenek düzeyinde kabul edilmektedir. Örneğin, 20 maddelik bir matematik testini alan ve farklı yedi maddeye doğru yanıt veren iki öğrenci olduğu varsayılırsa bu öğrencilerin ayırt edebilmesi sorunu ortaya çıkacaktır. Doğru yanıt verilen maddelerin güçlüklerinin ve ayırt ediciliklerinin önemli derecede birbirinden farklı olması durumunda iki bireyin aynı yetenek düzeyinde kabul edilmesinin ölçme sonuçlarının hatalı yorumlanmasına yol açacağı öngörülebilir. Böyle bir durumda, test sonuçlarını eğitimin iyileştirilmesi için bir dönüt olarak kullanacak olan öğretmenin yedi kolay soruya doğru yanıt veren öğrenci ile yedi zor soruya doğru yanıt veren öğrenci için uygulayacağı yöntemleri seçerken test puanlarını göz önünde bulundurması yetersiz olmanın ötesinde yanıltıcı da olacaktır. Özellikle biçimlendirici değerlendirme (formative assesment) durumlarında KTK’ya dayalı ölçme süreçlerinden istenilen düzeyde etkili sonuçlar elde edilemeyeceği görüşü öne sürülebilir. MTK modelleri farklı yanıt örüntüsüne sahip eşit sayıda doğru yanıt bulunan bireylerin yetenek düzeyi bakımından birbirlerinden ayırt edilebilmesi konusunda etkili çözümler sunmaktadır fakat MTK’ya dayalı parametrelerin daha hatasız kestirilebilmesi için uygulamaların geniş örneklem üzerinde yapılması gerekmektedir (Hulin, Lissak ve Drasgow, 1982; Ree ve Jensen, 1983; Hambleton ve van der Linden, 1997; de Ayala, 2009). Okullarda yapılan ölçmeler ise küçük katılımcı grupları üzerinde yürütülmekte bu durum başlı başına MTK’ya dayalı test uygulamalarını olanaksız kılmaktadır. Üstelik MTK’ya dayalı test geliştirmek ve sonuçları değerlendirmek alan uzmanlığı gerektiren bir süreçtir. Söz konusu nedenlerden dolayı öğretmen yapımı testlerin geliştirilmesi ve test sonuçlarının değerlendirilmesi sürecinde KTK’ya dayalı tekniklerin kullanılmasının daha uygun olduğu görüşü öne sürülebilir. Buna karşın KTK’nın sınırlılıkları düşünüldüğünde ölçme sonuçlarının geçerliği, güvenilirliği ve karşılaştırılabilirliği bakımından önemli sorunların ortaya çıkması küçük örneklemelerde daha etkili ölçmelerin yapılabilmesini sağlayacak farklı yöntem ve tekniklerin geliştirilmesi ihtiyacını doğurmaktadır.

Takahiro Sato 1969’da diğer test kuramlarının sınırlılıklarından doğan ihtiyaçtan hareketle katılımcıların cevap örüntüleri üzerinden performanslarını ayırt etmeye yönelik küçük örneklem üzerinde de etkili sonuçlar üretebilecek bir sistem üzerinde çalışmalarını yürütmüştür. Sato’nun geliştirmiş olduğu Öğrenci-Madde Çizelgesi Analizi (Student-Problem Chart Analysis) ile hem maddelerin kalitesi hem de öğrencilerin bilme ve öğrenme durumları belirlenebilmektedir. Bu anlamda Sato Test Kuramı modern test kuramları arasında yerini almaktadır (Switzer ve Connell, 1990; Wu, 1999; Sheu, Nguyen, Nguyen, Pham, Tsai ve Nagai, 2014).

Öğrenci-Madde Çizelgesi Analizi ile *Öğrenci Uyarı İndeksi* (Student Caution Index) ve *Madde Uyarı İndeksi* (Problem Caution Index) olmak üzere iki tür indeks üretilmektedir. Öğrenci Uyarı İndeksi (ÖUI) aracılığıyla öğrenciler sınıflandırılarak her bir öğrencinin testle ölçülen konu

alanlarını öğrenme ve maddeleri dikkatli bir şekilde yanıtlama durumlarına ilişkin dönüt alınmaktadır. Benzer şekilde Madde Uyarı İndeksi (MUI) aracılığıyla maddeler sınıflandırılarak her bir maddenin işlerliği hakkında dönüt alınmaktadır. Bu indeksler aşağıdaki formüllerle hesaplanır (Sheu, Nguyen, Ngyuyen, Pham, Tsai ve Nagai, 2014; Lin ve Yih, 2015).

$$\text{ÖÜİ} = 1 - \frac{\sum_{m=1}^M (Y_{nm})(Y_m) - (Y_n) \cdot (u')}{\sum_{m=1}^{Y_n} Y_m - (Y_n) \cdot (u')}$$

$$\text{MUI} = 1 - \frac{\sum_{n=1}^N (Y_{nm})(Y_n) - (Y_m) \cdot (\mu')}{\sum_{n=1}^{Y_m} Y_n - (Y_m) \cdot (\mu')}$$

m : Maddeler ($m=1,2,3,\dots,M$)

n : Öğrenciler ($n=1,2,3,\dots,N$)

Y_{nm} : n . öğrencinin m . maddeden aldığı puan (doğru:1, yanlış: 0)

Y_m : Öğrencilerin m . maddeden aldıkları puanların toplamı

Y_n : n . öğrencinin toplam test puanı

$$u' : \frac{\sum_{m=1}^M Y_m}{M}, \quad \mu' : \frac{\sum_{n=1}^N Y_n}{N},$$

Madde Uyarı İndeksi (MUI), maddeleri dört farklı sınıfa atamak için kullanılır. Her bir madde için hesaplanan MUI değerinin 0,50'den daha büyük bir değer alması ilgili maddede bazı sorunların olabileceğine ilişkin bilgi verir. Elde edilen MUI değeri maddenin doğru yanıtlanma oranı ile birlikte yorumlanır. Yapılan sınıflandırmalar ve bunlara ilişkin yapılabilecek yorumlar aşağıda verilmiştir (Sheu, Pham, Nguyen ve Nguyen, 2013).

A: Düzgün çalışan bir madde.

A': Kısmi olarak sorunlu bir madde. Tekrar gözden geçirilebilir.

B: Zor bir madde.

B': Sorunlu bir madde. Ters işliyor olabilir. Testten atılmalı ya da yeniden yazılmalı.

Öğrenci Uyarı İndeksi (ÖÜİ) ise, öğrencileri altı farklı sınıfa atamak için kullanılır. Her bir öğrenci için hesaplanan ÖÜİ değerinin 0,50'den daha büyük bir değer alması ilgili öğrencinin öğrenme eksikliklerinin olabileceğine ya da dikkatsizlikle kaçırıldığı soruların olabileceğine ilişkin bilgi verir. Elde edilen ÖÜİ değeri öğrencinin maddeleri doğru yanıtlama oranı ile birlikte yorumlanır. Yapılan sınıflandırmalar ve bunlara ilişkin yapılabilecek yorumlar aşağıda verilmiştir (Sheu, Pham, Nguyen ve Nguyen, 2013).

A: Etkili öğrenme gerçekleşmiş.

A': Öğrenme var fakat çok dikkatsiz.

B: Genel olarak iyi ancak biraz daha çalışmaya ihtiyacı var.

B': Biraz dikkatsiz ve daha fazla çalışmaya ihtiyacı var.

C: Öğrenme düzeyi yetersiz.

C': Öğrenme gerçekleşmemiş.

Söz konusu ÖÜİ ve MUI indekslerinin öğrencilerin yetenek düzeyleri ve maddelerin kalite düzeyleri hakkında yeterli bilgi vermediği iddia edilebilir. Bunun nedeni, nitel veri düzeyindeki çıktılarının, bireyleri ve maddeleri karşılaştırabilme, yapılan değerlendirme sonucu çeşitli kararlar alabilme anlamında sınırlı kalmasıdır. Öğrenci-Madde Çizelgesi'nin bu sınırlılığını ortadan kaldırmak

amacıyla Nagai (2010) bulanık mantığa (fuzzy logic) dayalı Gri İlişki Analizi'nin Sato Test Kuramı'na dahil edilmesini önererek Gri Öğrenci-Madde Çizelgesi Analizi'ni (Grey Student-Problem Chart Analysis) geliştirmiştir. Bu analize dayalı olarak öğrencilerin test performansları üzerinden üretilen *Öğrenciler için Sınırlandırılmış Gri İlişki Derecesi –ÖİSGİD-* (Localized Grey Relational Grade-Student) ve maddelere verilen yanıtlar üzerinden üretilen *Maddeler için Sınırlandırılmış Gri İlişki Derecesi –MİSGİD-* (Localized Grey Relational Grade-Problem) elde edilmektedir (Sheu, Nguyen, Ngyuyen, Pham, Tsai ve Nagai, 2014). Gri ilişki dereceleri hesaplanırken kullanılan vektörler ve formüller aşağıda verilmiştir (Sheu, Nguyen, Tsai, Pham, Ngyuen ve Nagai, 2014).

Gri ilişki analizinde karşılaştırılan vektörler,

$$x_0 = (x_0(1), x_0(2), \dots, x_0(k), \dots, x_0(n)) : \text{Referans vektör}$$

$$x_i = (x_i(1), x_i(2), \dots, x_i(k), \dots, x_i(n)) : \text{İncelenen vektör}$$

Gri öğrenci-madde çizelgesi analizinde incelenen vektörler,

$$S_i = (x_i(1), x_i(2), \dots, x_i(k), \dots, x_i(n)) : i. \text{ öğrencinin tüm maddelere verdiği yanıtlar vektörü, } k = 1, 2, 3, \dots, n$$

$$P_k = (x_1(k), x_2(k), \dots, x_i(k), \dots, x_m(k)) : k. \text{ maddeye tüm öğrencilerin verdiği yanıtlar vektörü, } i = 1, 2, 3, \dots, m$$

$$\text{ÖİSGİD} = \Gamma_{(x_0, x_i)} = \frac{\bar{\Delta}_{max} - \bar{\Delta}_{0i}}{\bar{\Delta}_{max} - \bar{\Delta}_{min}}$$

$$\text{MİSGİD} = \Gamma_{(x_j, x_0)} = \frac{\bar{\Delta}_{max} - \bar{\Delta}_{k0}}{\bar{\Delta}_{max} - \bar{\Delta}_{min}}$$

$$\bar{\Delta}_{0i} = \left(\sum_{k=1}^n (x_0(k) - x_i(k))^p \right)^{\frac{1}{p}} : \text{Referans vektör ve incelenen vektör (öğrenciler için) arasındaki Minkowski uzaklık ölçüsü}$$

$$\bar{\Delta}_{k0} = \left(\sum_{i=1}^m (x_0(k) - x_i(k))^p \right)^{\frac{1}{p}} : \text{Referans vektör ve incelenen vektör (maddeler için) arasındaki Minkowski uzaklık ölçüsü}$$

$$\bar{\Delta}_{max} = \bar{\Delta}_{0i} \text{ ve } \bar{\Delta}_{k0} \text{ için maksimum değer}$$

$$\bar{\Delta}_{min} = \bar{\Delta}_{0i} \text{ ve } \bar{\Delta}_{k0} \text{ için minimum değer}$$

Gri İlişki Analizi'nin dahil edilmesiyle birlikte Sato Test Kuramı öğrenciler ve öğretim sürecinin etkililiği ile ilgili dönüt alabilmek için kullanılabilmesinin (biçimlendirici değerlendirme) yanı sıra düzey belirlemeye ve buna bağlı olarak öğrencilerin eğitim durumları ile ilgili karar almaya (özetleyici değerlendirme) yönelik olarak kullanılabilir.

Sato Test Kuramı üzerinde Türkiye'de çok az sayıda çalışma yapıldığı ve kurama ilişkin uygulamaların genellikle Uzak Doğu Ülkelerinde yaygınlaştığı belirtilebilir. Günümüzde eğitim, psikoloji, sağlık ve endüstri gibi pek çok farklı alanda ölçme ve değerlendirme çalışmalarına verilen

önem yaygınlaşmakta ve ölçme faaliyetlerinin önemli bir bölümü küçük katılımcı grupları üzerinde gerçekleştirilmektedir. Bu bağlamda değerlendirildiğinde Sato Kuramcılarının küçük örneklem üzerinde etkili sonuçların elde edilmesine olanak sağlayan yöntemler sunduklarına dair iddialarının dikkate alınmaya değer olduğu sonucuna varılabilir. Kurama ilişkin yöntemlerin geçerliği ve güvenilirliğine ilişkin çalışmaların sayısının artmasının kuramın işlerliğinin ve kullanılabilirliğinin ortaya koyulması bakımından büyük önem taşıdığı açıktır.

Araştırmanın Amacı

Bu araştırmanın amacı, aynı evrenden tesadüfi olarak seçilen örneklemelerden elde edilen Maddeler için Sınırlandırılmış Gri İlişki Derecelerinin (MİSGİD) tutarlılığının ve Madde Uyarı İndeksleri (MUI) aracılığı ile gerçekleştirilen madde sınıflandırmaları arasındaki uyumun incelenmesidir. Bu amaç doğrultusunda aşağıdaki sorulara yanıt aranmıştır:

- 1) Örneklemelerden elde edilen Maddeler için Sınırlandırılmış Gri İlişki Dereceleri arasında ne düzeyde bir ilişki bulunmaktadır?
- 2) Örneklemelerden elde edilen Maddeler için Sınırlandırılmış Gri İlişki Derecelerinin ortalamaları arasındaki fark istatistiksel olarak anlamlı mıdır?
- 3) Örneklemelerden elde edilen Madde Uyarı İndeksleri ile yapılan sınıflandırmalar ne oranda uyum göstermektedir?

YÖNTEM

Bu çalışmada, simülatif olarak oluşturulmuş 20 maddelik bir testi yanıtlayan 1000 kişi içerisinde seçilen 30'ar kişilik 10 farklı örneklemelerden elde edilen Maddeler için Sınırlandırılmış Gri İlişki Derecelerinin (MİSGİD) tutarlılığının ve Madde Uyarı İndeksleri (MUI) aracılığı ile gerçekleştirilen madde sınıflandırmaları arasındaki uyumun incelenmesi amaçlanmıştır. Araştırma, bilgi üretmeye yönelik kuramsal bir çalışma özelliği taşıması bakımından temel araştırma niteliğindedir.

Verilerin Oluşturulması

Araştırmada, WinGen3 programı ile yetenek düzeyleri -3 ve +3 aralığında değişen 1000 kişilik bir katılımcı grubu ile a parametresi 0 ve 2 aralığında; b parametresi -3 ve +3 aralığında; c parametresi 0 ve 0,20 aralığında değişen 20 maddeden oluşan, 1-0 şeklinde puanlanan bir test oluşturularak her bir katılımcı için yanıt örüntüsü elde edilmiştir. WinGen programı genellikle Madde Tepki Kuramı kapsamında yapılan çalışmalar için simülasyon veri üretmek amacıyla kullanılmaktadır. Bu nedenle maddeler MTK parametrelerine dayalı olarak üretilmiş olmasına karşın bu çalışma için parametrelerin çok fazla öneminin olmadığı belirtilebilir. Araştırmada WinGen programı gerçeğe daha uygun yanıt örüntüleri elde etmek amacıyla tercih edilmiştir.

Yanıt örüntüleri elde edildikten sonra bunların içerisinde seçkisiz olarak seçilen ve 30'ar kişinin yanıt örüntülerinden oluşan 10 farklı veri seti (örneklem) oluşturulmuştur. Söz konusu örneklemeler ve maddelerin daha iyi betimlenebilmesi amacıyla her bir örnekleme seçilen yanıt örüntüleri üzerinden Klasik Test Kuramı yaklaşımıyla madde güçlük indeksleri (p , doğru yanıt puan toplamı/30) ve madde ayırt edicilik indeksleri (r , madde-toplam test puanları arası Çift Serili Korelasyon Katsayısı) hesaplanmıştır. Elde edilen bulgular Tablo 1 ve Tablo 2 de sunulmuştur.

Tablo 1. Farklı örneklerden elde edilen madde güçlük indeksleri

Maddeler	Örn1	Örn2	Örn3	Örn4	Örn5	Örn6	Örn7	Örn8	Örn9	Örn10
	p	p	p	p	p	p	p	p	p	p
1	0,13	0,13	0,10	0,13	0,07	0,23	0,07	0,27	0,17	0,20
2	0,50	0,67	0,50	0,43	0,50	0,57	0,47	0,53	0,40	0,50
3	0,30	0,40	0,37	0,40	0,40	0,57	0,17	0,37	0,30	0,57
4	0,70	0,53	0,67	0,80	0,60	0,77	0,60	0,67	0,70	0,67
5	0,40	0,23	0,30	0,43	0,27	0,43	0,43	0,40	0,33	0,30
6	0,43	0,47	0,40	0,50	0,47	0,60	0,40	0,50	0,60	0,57
7	0,43	0,37	0,33	0,63	0,47	0,47	0,40	0,43	0,40	0,47
8	0,83	0,87	0,83	0,90	0,77	0,90	0,97	0,77	0,97	0,87
9	0,40	0,33	0,37	0,57	0,40	0,57	0,43	0,60	0,43	0,50
10	0,30	0,43	0,30	0,43	0,37	0,60	0,30	0,43	0,50	0,57
11	0,90	0,87	0,80	0,87	0,80	0,87	0,77	0,77	0,87	0,80
12	0,33	0,23	0,30	0,53	0,37	0,53	0,33	0,50	0,37	0,40
13	0,27	0,23	0,23	0,30	0,23	0,40	0,27	0,40	0,23	0,23
14	0,87	0,87	0,87	0,90	0,97	0,90	0,90	0,93	0,87	0,90
15	0,23	0,27	0,20	0,37	0,23	0,30	0,37	0,30	0,30	0,33
16	0,73	0,53	0,57	0,77	0,57	0,70	0,57	0,67	0,73	0,70
17	0,40	0,50	0,43	0,50	0,43	0,63	0,33	0,67	0,47	0,50
18	0,33	0,17	0,20	0,17	0,17	0,23	0,27	0,17	0,23	0,13
19	0,50	0,70	0,57	0,60	0,67	0,77	0,57	0,37	0,63	0,63
20	0,47	0,50	0,53	0,47	0,40	0,63	0,50	0,47	0,53	0,50

Tablo 1 incelendiğinde, 1, 13, 15 ve 18. maddelerin zor; 8,11 ve 14. maddelerin ise kolay maddeler olduğu görüşü öne sürülebilir.

Tablo 2. Farklı örneklerden elde edilen madde ayırt edicilik indeksleri

Maddeler	Örn1	Örn2	Örn3	Örn4	Örn5	Örn6	Örn7	Örn8	Örn9	Örn10
	r	r	r	r	r	r	r	r	r	r
1	0,21	0,26	0,07	0,25	0,24	0,02	0,06	0,30	0,13	0,05
2	0,33	0,18	0,18	0,43	0,21	0,44	0,24	0,52	0,33	0,68
3	0,28	0,50	0,66	0,65	0,40	0,01	0,56	0,53	0,30	0,37
4	0,56	0,68	0,56	0,46	0,70	0,52	0,79	0,69	0,55	0,67
5	0,56	0,65	0,48	0,51	0,34	0,60	0,64	0,51	0,43	0,57
6	0,53	0,58	0,70	0,59	0,68	0,06	0,62	0,71	0,65	0,86
7	0,53	0,79	0,77	0,76	0,72	0,64	0,64	0,64	0,77	0,72
8	0,64	0,50	0,50	0,38	0,58	0,29	0,21	0,64	0,28	0,52
9	0,42	0,28	0,43	0,51	0,62	0,01	0,68	0,52	0,40	0,68
10	0,41	0,41	0,38	0,58	0,74	0,17	0,57	0,31	0,53	0,37
11	0,41	0,35	0,50	0,37	0,34	0,08	0,58	0,32	0,31	0,55
12	0,36	0,28	0,28	0,41	0,45	0,40	0,44	0,53	0,22	0,27
13	0,43	0,26	0,60	0,50	0,16	0,60	0,21	0,62	0,46	0,41
14	0,62	0,33	0,41	0,43	0,13	0,40	0,41	0,21	0,31	0,47
15	0,53	0,57	0,38	0,49	0,40	0,73	0,33	0,57	0,46	0,09
16	0,51	0,75	0,62	0,71	0,79	0,16	0,78	0,74	0,41	0,67
17	0,31	0,52	0,31	0,40	0,46	0,10	0,31	0,30	0,33	0,23
18	0,53	0,40	0,35	0,33	0,24	0,66	0,04	0,34	0,31	0,35
19	0,21	0,39	0,28	0,05	0,26	0,48	0,04	0,48	0,04	0,06
20	0,62	0,26	0,43	0,24	0,36	0,03	0,03	0,17	0,26	0,10

Tablo 2 incelendiğinde, Çift Serili Korelasyon Katsayısı yöntemi ile hesaplanan madde ayırt edicilik indekslerinin farklı örneklem arasında büyük farklılıklar gösterdiği görülmektedir. Örneğin, 3. maddenin ayırt edicilik indeksi Örneklem 3 verileri ile hesaplandığında 0,66 değerini alırken Örneklem 6 verileri ile hesaplandığında 0,01 değerini almaktadır. Benzer şekilde 20. maddenin ayırt edicilik indeksi Örneklem 1’de 0,62 olarak hesaplanmış fakat Örneklem 6 ve Örneklem 7 de 0,03 olarak belirlenmiştir. Tablodan hareketle birçok madde için benzer durumlardan söz edilebilir. Klasik Test Kuramı yöntemleri küçük örneklemelerde madde parametreleri hesaplamak için sıklıkla kullanılmaktadır fakat karşılaşılan bu tip durumlar KTK’nın bir dezavantajı olarak ifade edilebilir.

Verilerin Analizi

Çalışmada, Maddeler için Sınırlandırılmış Gri İlişki Dereceleri (MİSGİD) seçkisiz olarak belirlenen 10 farklı örneklemin her birinden, Sheu, Pham, Nguyen ve Nguyen (2013) tarafından Öğrenci-Madde Çizelgesi Analizi ve Gri Öğrenci-Madde Çizelgesi Analizi için geliştirilen Matlab Paketi ile elde edilmiştir. Paket, Matlab R2013b versiyonu üzerinde çalıştırılmıştır. Elde edilen MİSGİD değerlerinin farklı örneklemler arası tutarlılığının belirlenebilmesi amacıyla Spearman Sıra Farkları Korelasyon Katsayıları hesaplanmış ve buna ek olarak değer ortalamaları arasındaki farkın anlamlılığı Kruskal Wallis-H Testi ile incelenmiştir. Söz konusu analizler için SPSS V22 programı kullanılmıştır. Son olarak örneklemelerden elde edilen Madde Uyarı İndeksleri ile yapılan sınıflandırmaların ne oranda uyum gösterdiği yüzde oranlar yardımıyla incelenmiştir.

BULGULAR

Öncelikle, “Örneklemelerden elde edilen Maddeler için Sınırlandırılmış Gri İlişki Dereceleri (MİSGİD) arasında ne düzeyde bir ilişki bulunmaktadır?” sorusuna yanıt bulmak amacıyla, 10 farklı veri setinden (örneklem) elde edilen MİSGİD değerleri hesaplanmış ve elde edilen bulgular Tablo 3’te sunulmuştur.

Tablo 3. Farklı Örneklemelerden Elde Edilen MİSGİD Değerleri

Maddeler	Örn1	Örn2	Örn3	Örn4	Örn5	Örn6	Örn7	Örn8	Örn9	Örn10
1	0,07	0,07	0,05	0,07	0,12	0,03	0,03	0,14	0,09	0,11
2	0,42	0,29	0,29	0,25	0,34	0,29	0,27	0,32	0,23	0,32
3	0,23	0,16	0,20	0,23	0,34	0,23	0,09	0,20	0,16	0,16
4	0,32	0,45	0,42	0,55	0,52	0,37	0,37	0,42	0,45	0,42
5	0,12	0,23	0,16	0,25	0,25	0,14	0,25	0,23	0,18	0,16
6	0,27	0,25	0,23	0,29	0,37	0,27	0,23	0,29	0,37	0,34
7	0,20	0,25	0,18	0,39	0,27	0,27	0,23	0,25	0,23	0,27
8	0,63	0,59	0,59	0,68	1,00	0,52	0,82	0,52	0,82	0,63
9	0,18	0,23	0,20	0,34	0,34	0,23	0,25	0,37	0,25	0,29
10	0,25	0,16	0,16	0,25	0,37	0,20	0,16	0,25	0,29	0,34
11	0,63	0,68	0,55	0,63	0,63	0,55	0,52	0,52	0,63	0,55
12	0,12	0,18	0,16	0,32	0,32	0,20	0,18	0,29	0,20	0,23
13	0,12	0,14	0,12	0,16	0,23	0,12	0,14	0,23	0,12	0,12
14	0,63	0,63	0,63	0,68	0,68	0,82	0,68	0,74	0,63	0,68
15	0,14	0,12	0,11	0,20	0,16	0,12	0,20	0,16	0,16	0,18
16	0,32	0,48	0,34	0,52	0,45	0,34	0,34	0,42	0,48	0,45
17	0,29	0,23	0,25	0,29	0,39	0,25	0,18	0,42	0,27	0,29
18	0,09	0,18	0,11	0,09	0,12	0,09	0,14	0,09	0,12	0,07
19	0,45	0,29	0,34	0,37	0,52	0,42	0,34	0,20	0,39	0,39
20	0,29	0,27	0,32	0,27	0,39	0,23	0,29	0,27	0,32	0,29

Tablo 3 incelendiğinde, 10 farklı örneklem üzerinden yapılan hesaplamalarda 8. madde haricinde diğer tüm maddelerin 0,75'in altında MİSGİD değeri elde ettiği görülmektedir. MİSGİD değerinin 0-1 aralığında değiştiği göz önünde bulundurulduğunda madde kalitelerinin yüksek olmadığı belirtilebilir. Daha sonraki aşamada, elde edilen MİSGİD değerleri arasındaki ilişkiler Spearman Sıra Farkları Korelasyon Katsayısı tekniği ile belirlenmiş ve elde edilen bulgular Tablo 4'te sunulmuştur.

Tablo 4. MİSGİD Değerleri Arasındaki Korelasyonlara İlişkin Bulgular

	Örn1	Örn2	Örn3	Örn4	Örn5	Örn6	Örn7	Örn8	Örn9	Örn10
Örn1										
Örn2	,874*									
Örn3	,950*	,933*								
Örn4	,770*	,866*	,857*							
Örn5	,934*	,841*	,954*	,844*						
Örn6	,943*	,926*	,956*	,893*	,904*					
Örn7	,829*	,929*	,879*	,852*	,809*	,852*				
Örn8	,741*	,780*	,813*	,832*	,793*	,781*	,733*			
Örn9	,899*	,886*	,919*	,895*	,959*	,899*	,873*	,814*		
Örn10	,922*	,858*	,900*	,867*	,935*	,912*	,853*	,819*	,972*	

*Korelasyon ,01 düzeyinde anlamlıdır.

Tablo 4 incelendiğinde, farklı örneklemelerden 20 madde için elde edilen MİSGİD değerlerinin birbirleri ile yüksek düzeyde pozitif korelasyon gösterdiği ve söz konusu korelasyon değerlerinin ,01 düzeyinde anlamlı olduğu görülmektedir. Bu bulgudan hareketle, Sato Test Kuramı kapsamında hesaplanan ve maddelerin kalitesi hakkında bilgi veren Maddeler için Sınırlandırılmış Gri İlişki Derecelerinin, aynı maddeleri yanıtlayan katılımcılar arasından seçilen farklı örneklemelerde tutarlı sonuçlar verdiği iddia edilebilir. Bu iddiayı sağlamlaştırabilmek amacıyla, “Örneklemelerden elde edilen Maddeler için Sınırlandırılmış Gri İlişki Derecelerinin ortalamaları arasındaki fark istatistiksel olarak anlamlı mıdır?” sorusuna yanıt aranmış ve farklı örneklemelerden elde edilen MİSGİD değerlerinin ortalamaları arasındaki farkın anlamlılığı nonparametrik bir test olan Kruskal Wallis-H testi ile incelenmiştir. Elde edilen bulgular Tablo 5'te sunulmuştur.

Tablo 5. MİSGİD Değerlerinin Ortalamaları Arasındaki Farkın Anlamlılığına İlişkin Bulgular

	MİSGİD
Ki-Kare	8,929
sd	9
p	,444

Tablo 5 incelendiğinde, MİSGİD değerlerinin ortalamaları arasındaki farkın istatistiksel olarak anlamlı olmadığı ($p=0,444$) görülmektedir. Bu bulgunun, farklı örneklemelerden elde edilen MİSGİD parametrelerinin tutarlı sonuçlar verdiği ilişkin ikincil bir kanıt oluşturduğu ifade edilebilir.

Son olarak, “Örneklemelerden elde edilen Madde Uyarı İndeksleri (MUI) ile yapılan sınıflandırmalar ne oranda uyum göstermektedir?” sorusuna yanıt aranmış, yapılan sınıflandırmalar ve bunlara ilişkin uyum yüzdeleri Tablo 6'da sunulmuştur.

Tablo 6. Madde Uyarı İndeksleri ile Yapılan Sınıflandırmaların Uyumuna İlişkin Bulgular

Maddeler	Örn1	Örn2	Örn3	Örn4	Örn5	Örn6	Örn7	Örn8	Örn9	Örn10	Uyum
1	B	B	B'	B	B'	B	B'	B'	B'	B'	%60
2	A'	A'	A'	B'	A'	A'	B'	A'	B'	A'	%70
3	B	B'	B	B	A'	B'	B	B	B'	B	%60
4	A	A	A	A	A	A	A	A	A	A	%100
5	B	B	B	B	B	B	B	B	B	B	%100
6	B	B	B	A	A	B	B	B	A	A	%60
7	B	B	B	A	B	B	B	B	B	B	%90
8	A	A	A	A	A	A	A	A	A	A	%100
9	B'	B	B	A	A'	B	B	A	B	A	%50
10	B	B	B	B	A'	B	B	B'	A	A'	%60
11	A	A	A	A	A	A	A	A'	A	A	%90
12	B'	B'	B'	A	A	B'	B	A	B'	B'	%60
13	B'	B	B	B	B	B'	B'	B	B'	B	%60
14	A	A	A	A	A	A	A	A	A	A	%100
15	B	B	B	B	B	B	B'	B	B	B'	%80
16	A	A	A	A	A	A	A	A	A	A	%100
17	A	B'	B'	A	A'	B	B'	A'	B'	A'	%40
18	B	B	B	B	B	B'	B'	B'	B'	B	%60
19	A'	A'	A'	A'	A'	A'	A'	B'	A'	A'	%90
20	A'	B	A	B'	A'	B'	A'	B'	A'	A'	%50

Tablo 6'da yer alan sınıflandırmaların farklı örneklemeler üzerinden yapılan hesaplamalar sonucu değişkenlik gösterebildiği belirlenmiştir. Buna karşın 4, 5, 8, 14 ve 16. maddelerin 10 farklı örnekleme de aynı şekilde (%100 uyum) sınıflandırıldığı, 2, 7, 11, 15 ve 19. maddelerin de sınıflandırma uyumlarının yüksek olduğu belirlenmiştir (uyum yüzdeleri için modu en yüksek olan sınıf baz alınmıştır). Sınıflandırılma bakımından en fazla değişkenlik gösteren maddenin 17. madde olduğu (%40 uyum) ifade edilebilir. Söz konusu madde haricindeki tüm maddelerin %50 ve üzeri uyum oranında sınıflandırıldığı bulgusu göz önünde bulundurulduğunda Madde Uyarı İndekslerine dayalı olarak yapılan sınıflandırmaların tutarlı olduğu görüşü öne sürülebilir.

SONUÇLAR ve TARTIŞMA

Bu araştırmada, simülatif olarak oluşturulmuş 20 maddelik bir testi yanıtlayan 1000 kişi içerisinde seçilen 30'ar kişilik on farklı örneklemden elde edilen Maddeler için Sınırlanmış Gri İlişki Derecelerinin tutarlılığının ve Madde Uyarı İndeksleri aracılığı ile gerçekleştirilen madde sınıflandırmaları arasındaki uyumun incelenmesi amaçlanmıştır.

Farklı örneklemelerden 20 madde için elde edilen MİSGİD değerlerinin birbirleri ile yüksek düzeyde pozitif korelasyon gösterdiği ve söz konusu korelasyon değerlerinin ,01 düzeyinde anlamlı olduğu belirlenmiştir. Buna ek olarak Kruskal Wallis-H testi ile yapılan incelemelerde MİSGİD değerlerinin ortalamaları arasındaki farkın istatistiksel olarak anlamlı olmadığı ($p= 0,444$) görülmüştür. Elde edilen bulgulardan hareketle, Sato Test Kuramında maddelerin kalitesinin belirlenebilmesi için üretilen MİSGİD parametrelerinin oldukça küçük örneklemeler (30 kişi) üzerinden hesaplanmasına karşın tutarlı ve yakın değerler aldığı sonucuna ulaşılmıştır.

Madde Uyarı İndeksleri ile yapılan sınıflandırmaların örneklemeler arası değişkenlik gösterdiği belirlenmiş buna karşın beşi %100 uyum göstermekle birlikte toplamda 10 maddenin yüksek düzeyde uyum oranıyla sınıflandırıldığı ve bir madde haricinde %50'nin altında uyum oranıyla sınıflandırılan madde olmadığı görülmüştür. Elde edilen bulgular, Sato Test Kuramı kapsamında yapılan test analizlerinin küçük örneklemelerde bile iyi sonuçlar verdiği iddialarına kanıt

oluşturabilecek niteliktedir. Sheu ve diğerleri (2014), Gri Öğrenci-Problem Çizelgesi Analizinin yalnızca küçük örneklemelerde değil büyük örneklemelerde de kullanılabileceği sonucuna ulaşmışlardır. Bu çalışma ise, küçük örneklemeler üzerinde Sato Test Kuramına dayalı üretilen madde parametrelerinin tutarlılığının incelenmesi süreciyle sınırlandırılmıştır. Alanyazında benzer çalışmalara ulaşılamaması bu konu üzerinde yapılabilecek tartışmaları sınırlandırmaktadır. Sato Test Kuramıyla ilgili çalışmaların son yıllarda yalnızca Uzak Doğu Ülkelerinde bir artış gösterdiği bunların da kuramı tanıtmaya (Lin ve Chen, 2006; Wang, Sheu ve Nagai, 2011; Wang ve Chen 2013) ya da kurama ilişkin geliştirilen bir yazılımı tanıtmaya (Wu, 1999; Sheu, Pham, Nguyen ve Nguyen, 2013; Sheu, Nguyen, Nguyen, Pham, Tsai ve Nagai, 2014; Pham, Sheu ve Nagai, 2015) amacıyla yapıldığı belirlenmiştir. Bu anlamda, kuram çerçevesinde üretilen parametrelerin çeşitli koşullar altında incelenmesi ve bu alanda yapılan çalışmaların sayısının artırılması gerekliliği ortaya çıkmaktadır. Benzer çalışmaların sayısının artması ve kuramın güvenilir ve geçerli sonuçlar üretebildiğine dair bulguların fazlalaşması, küçük örneklemelerde Klasik Test Kuramına alternatif yöntemlerin kullanımının yaygınlaşması bakımından oldukça önemlidir. Söz konusu çabalar, daha etkili ölçme araçları geliştirme, bireylerin bir takım özellikleriyle ilgili daha doğru bilgiler elde edebilme ve bireyler hakkında yapılan değerlendirmelerin isabetliliği bakımından alandaki ilerlemeye katkı sağlayacaktır. Tüm bunların yanısıra, kuramın uygulanabilirliği, iç geçerliliğin sağlanmasıyla ilişkili olduğu kadar dış geçerliliğin sağlanmasını da zorunlu kılmaktadır. Söz konusu nedenle, kurama ilişkin parametrelerin hem kendi içerisinde tutarlılığının belirlenebilmesi hem de diğer test kuramlarına dayalı yöntemlerle üretilen parametrelerle karşılaştırılarak sonuçların incelenmesi, daha sonra yapılacak çalışmalarda ele alınması bakımından önerilebilir.

KAYNAKÇA

- de Ayala, R. J. (2009). The theory and practice of item response theory. New York: The Guilford Press.
- Hambleton, R. K., Swaminathan, H., ve Rogers, H. (1991). Fundamentals of item response theory. Newbury Park CA: Sage.
- Hambleton, R. K., ve van der Linden, W. J. (1997). Handbook of modern item response theory. New York:Springer.
- Hulin, C. L., Lissak, R. I., ve Drasgow, F. (1982). Recovery of two and three-parameter logistic item characteristic curves: A Monte Carlo study. Applied Psychological Measurement, 6, 249-260.
- Lin, Y.H., ve Chen, S.M. (2006). The investigation of S-P chart analysis on the test evaluations of equality axiom concepts for sixth graders. Proceedings of the 2nd International Conference on Educational Technologies, Romania, Bucharest.
- Lin, Y.H., ve Yih, J.M. (2015). Application of IIRS in mathematics instruction to promote pupils decimal concept. The International Conference on Language, Education and Psychology, Taiwan.
- Pham, D.H., Sheu, T.W., ve Nagai, M. (2015). PCSP 1.0 software for partial credit S-P chart analysis. International Journal of Hybrid Information Technology. 8(6), 309-322.
- Ree, M. J., ve Jensen, H. E. (1983). Effects of sample size on linear equating of item characteristic curve parameters. In D. J. Weiss & R. D. Bock (Eds.), New horizons in testing: Latent trait test theory and computerized adaptive testing (pp. 135-146). New York: Academic Press.
- Suen, H.K. (1990). Principles of test theories. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Sheu, T. W., Nguyen, P. T., Nguyen, P. H., Pham, D. H., Tsai, P. C., ve Nagai, M. (2014). A MATLAB toolbox for misconceptions analysis based on S-P chart, grey relational analysis and ROC. Transactions on Machine Learning and Artificial Intelligence, 2, 72-85.
- Sheu, T.W., Nguyen, P.T., Tsai, C.P., Pham, D.H., Nguyen, P.H. ve Nagai, M. (2014). Using grey student-problem chart in the evaluation of tests with large data sets. Education Practise and Innovation, 1(2), 2372-3106.
- Sheu, T.W., Pham, D.H., Nguyen, P.T., ve Nguyen, P.H. (2013). Amatlab toolbox for student-problem chart and grey student-problem chart and its application. International Journal of Kansei, 4(2), 75-86.
- Switzer, D. M. ve Connell, M. L. (1990). Practical applications of student response analysis, Educational Measurement: Issue and Practice, 9, 15-17.
- Wang, B.T., Sheu, T.W., ve Nagai, M. (2011). Evaluating the english-learning of engineering students using the grey S-P chart: a facebook case study in Taiwan. Global Journal of Engineering Education, 13(2), 51-56.

Wang, C.H. ve Chen, C.P. (2013). Employing online S-P diagnostic table for qualitative comments on test results. *The Electronic Journal of e-Learning*, 11(3), 263-271.

Wu, H. (1999). Software Based on S-P Chart Analysis and Its Applications. *Proceedings of the National Science Council*, 8, 102-107.

EXTENDED ABSTRACT

Introduction

Psychometrists search response for questions such as how test items will be improved, how they will be chosen, how given answers will be turn into numbers, how they will be commented and how quality of obtained data will be assessed, as depending on mathematical and statistical theories. Aforementioned theories are called "test theory" or "psychometric theory" (Suen, 1990).

Classical Test Theory (CTT) and Item Response Theory (IRT) have become prominent as main two test theories that they have been studied mostly from past up to the day. On the literature, it has been seen that some of strengthening and weaknesses of Classical Test Theory and Item Reaction Theory have been discussed. Applications require to be made on large samples in the order that parameters based on IRT are to be estimated more faultlessly (Hulin, Lissak and Drasgow, 1982; Ree and Jensen, 1983; Hambleton and van der Linden, 1997; de Ayala, 2009). Measurements made at schools are conducted on small participant groups, this case precludes test applications based on IRT at there. Furthermore, developing test based on IRT and assessing results are a process which requires domain expertise. Due to aforementioned reasons, a view has been suggested about that using techniques based on CTT is more appropriate for development of teacher - made tests and assessment of test results, but still limitations of CTT are considered, that important problems occur in terms of measurement results validity, reliability and comparability. Because of this there is a need to develop different method and techniques which will provide to make more efficient measurements on small samples. When it is assessed with this sense, it can be concluded that suggestions on which Sato Theorists present methods providing opportunity to obtain effective results on small samples are things to consider.

The goal of this research is to the examination of consistency of Localized Grey Relational Grade-Problem (LGRG-P) values and accommodation item classifications which are made via Problem Caution Index (PCI) between different samples that are randomly chosen from the same universe. In accordance with this purpose, the following questions searched for answers:

- 1) What a level of the relation is there among Localized Grey Relational Grade Problem obtained from Samples?
- 2) Is the difference among averages of Localized Grey Relational Grade Problem obtained from samples significant as statistic?
- 3) At what rate classifications made by Problem Caution Indexes obtained from samples show accommodation?

Method

The research has a basic research quality in terms of having feature to be a theoretic study related to produce information. On this research, 1000 response patterns for each participants that their ability levels change at the range -3 and +3 with consists of 20 items (dichotomously) which change at the range of -3 and +3 for "b" parameters; at the range of 0 and 0,20 for "c" parameters; at the range of 0 and 2 for "a" parameter on WinGen3 program. After response patterns have been obtained, 10 different dataset (sample) which are randomly chosen among them and consist of response patterns of 30 persons.

On the study, Used Matlab Pack which is developed for Analysis of Student-Problem Chart and Analysis of Grey Student-Problem Chart by Sheu, Pham, Nguyen and Nguyen (2013) to obtain that

Localized Grey Relational Grades. The pack has been run on Matlab R2013b version. In order to examination of consistency of obtained LGRG-P values among different samples, Spearman Rank-Order Correlation Coefficients have been calculated and additionally significance of difference among value averages has been examined by Kruskal Wallis-H Test via SPSSV22.

Results and Discussion

It has been determined that LGRG-P values obtained for 20 items from different samples show highly positive correlation with each other and aforementioned correlation values are significant at 0.01 level. Additionally, on the examinations which are made by Kruskal Wallis-H Test, it has been seen that the difference among averages of LGRG-P values are not statistically significant ($P=0,444$). From this point of obtained findings, although LGRG-P parameters produced to determine quality of items on Sato Test Theory are calculated through quite small (30 persons) samples, it has been concluded that consistent and close values are gotten.

It has been determined that Classifications made by Problem Caution Index have showed change among samples but it has been seen that five of them have showed 100% accommodation and total 10 items have been classified with high level accommodation ratio and except from one item, there is not any items classified with accommodation ratio under 50%.

Findings which are obtained have quality to provide evidence against suggestions that test analyses made within the concept of Sato Test Theory have given well-results even on small samples. This study has been limited with examination process of item parameters consistency produced only as depending on Sato Test Theory. With this sense, a necessity reveals about that parameters produced for individuals responding test within the frame of the theory are to be examined under various conditions. It can be emphasized on a view that many studies which are conducted under different conditions are required in the order that consistency of parameters related to the theory in itself and also accuracy of results is to be discussed as comparing with parameters produced by methods based on other test theories.