



---

---

## Likert Tipi Ölçeklerde Klasik Test Kuramı ile Rasch Analizinden Elde Edilen Yetenek Kestirimleri Arasındaki Uyumun Test Edilmesi

Mustafa İLHAN<sup>1</sup>

Neşe GÜLER<sup>2</sup>

---

---

*Geliş Tarihi: 2017-02-03*

*Kabul Tarihi: 2017-02-28*

### Öz

Bu araştırmada, Likert tipi ölçeklerde klasik test kuramı (KTK) ile Rasch analizinden elde edilen yetenek kestirimleri göreceli uyum, mutlak uyum ve dağılım özellikleri açısından karşılaştırılmıştır. Araştırmanın verileri; 200 öğretmenden oluşan bir çalışma grubuna, 10 madde içeren ve tek faktörlü bir yapıya sahip olan “sınıftaki farklılıklar için öğretmen yeterlikleri” ölçeğinin uygulanması ile toplanmıştır. Çalışmada, KTK ile Rasch modelinde hesaplanan yetenek ölçüleri arasındaki korelasyon katsayısı .95 olarak bulunmuş ve iki kuramla kestirilen yetenek ölçüleri arasında yüksek bir göreceli uyum olduğu sonucuna varılmıştır. İlişkili örneklem *t*-testinden elde edilen bulgular, KTK’da hesaplanan yetenek kestirimlerinin Rasch analizinde ulaşılan yetenek kestirimlerine kıyasla anlamlı derece daha yüksek olduğunu ortaya koymuş ve bu sonuç iki kurama göre kestirilen yetenek puanları arasında mutlak bir uyum bulunmadığına yönelik bir kanıt olarak değerlendirilmiştir. Son olarak, Rasch analizinde kestirilen yetenek puanlarının KTK’den elde edilen yetenek puanlarına kıyasla daha simetrik ve daha sivri bir dağılıma sahip olduğu saptanmıştır.

*Anahtar Kelimeler:* Likert tipi ölçek, klasik test kuramı, Rasch analizi, madde tepki kuramı

---

<sup>1</sup> Yrd. Doç. Dr., Dicle Üniversitesi, Ziya Gökalp Eğitim Fakültesi, [mustafailhan21@gmail.com](mailto:mustafailhan21@gmail.com)

<sup>2</sup> Doç. Dr., Sakarya Üniversitesi, Eğitim Fakültesi, [gnguler@gmail.com](mailto:gnguler@gmail.com)



---

---

## Testing the Agreement Between Ability Estimations Made Through Classical Test Theory and the Ones Made Through Rasch Analysis in Likert Type Scales

---

---

*Submitted by 2017-02-03*

*Accepted by 2017-02-28*

### Abstract

The ability estimations calculated in Classical Test Theory (CTT) with the ones calculated in Rasch analysis was compared in the present study in terms of *i*) relative agreement, *ii*) absolute agreement, and *iii*) distribution properties. The research data were collected from a study group of 200 teachers by means of Teacher Efficacy Scale for Classroom Diversity (TESCD) containing 10 items with one-factor structure. Correlation coefficient between ability measures in CTT and in Rasch model was found to be .95, and it was concluded that there was a high relative agreement between ability estimations made in these two theories. The findings obtained in paired samples *t*-test suggested that the ability estimations made in CTT were significantly higher than those made in Rasch analysis, and consequently, the result was considered as evidence that there were no absolute agreements between the ability scores estimated in these two theories. Finally, it was found that the ability scores estimated in Rasch analysis had more symmetrical and sharper distribution than the ones in CTT.

*Keywords:* Likert type scale, classical test theory, Rasch analysis, item response theory

## Giriş

Ölçme ve değerlendirme bilim dalının tarihsel gelişimi incelendiğinde Klasik Test Kuramı (KTK) ve Madde Tepki Kuramı (MTK) olarak adlandırılan iki temel test kuramı karşımıza çıkmaktadır. Eğitimde ve psikolojide; bireylerin yetenek düzeylerinin hesaplanmasında, madde ve test istatistiklerinin kestirilmesinde bu iki kuramdan yararlanılmaktadır (Kan, 2006). KTK, birçok araştırmacının daha aşına olduğu bir kuram olup 1904 yılında Sperman tarafından ortaya atılmıştır. Kuramın, gözlenen puanın gerçek puan ile hata puanı bileşenlerinden oluştuğu ( $X = T + E$ ) şeklindeki temel önermesi ise 1966 yılında Novick tarafından kavramsallaştırılmıştır (Van der Linden ve Hambleton, 1997). KTK'ya göre bireylerin yetenek düzeyleri, ölçme aracının her bir maddesinden aldığı puanların toplanmasıyla elde edilmektedir (Hwang, 2002). Dolayısıyla KTK'da ölçme aracındaki farklı maddelerin bireyin yetenek düzeyi hakkında yapılan kestirime katkısı eşit olmaktadır (Anshel, Weatherby, Kang ve Watson, 2009). KTK, gerektirdiği varsayımlarının karşılanabilmesinin (Hambleton ve Jones, 1993) ve matematiksel olarak hesaplanmasının kolay oluşu (Haiyang, 2010) sebebiyle uzun yıllar boyunca psikolojik ölçme araçlarının geliştirilmesi ve analizinde yaygın olarak kullanılmıştır (Hambleton, Swaminathan ve Rogers, 1991). Ancak KTK'da bireylerin yetenek düzeylerinin madde parametrelerine ve madde parametrelerinin de bireylerin yetenek düzeylerine bağlı olması, araştırmacıları KTK'ya ilişkin bu sınırlılığın aşılmasına imkân tanıyacak yeni bir kuram arayışına yöneltmiştir (Çıkrıkçı Demirtaşlı, 1996). Bu arayış neticesinde, 1930'lu yılların sonunda örtük özellikler teorisi olarak da isimlendirilen MTK geliştirilmiştir (Doğan ve Tezbaşaran, 2003).

MTK'ya göre bireylerin belli bir alandaki doğrudan gözlenemeyen yetenekleri ile bu alanı yoklayan test maddelerine verdikleri cevaplar arasında bir ilişki vardır ve bu ilişkinin matematiksel olarak ifade edilmesi mümkündür (Kelecioğlu, 2001). MTK'da, bir maddenin parametreleri maddeyi yanıtlayan bireylerin yetenek düzeyinden bağımsız olarak hesaplanabilmekte ve bireylerin yetenek düzeyleri de yanıtladığı madde örnekleminde bağımsız olarak kestirilebilmektedir (Embretson ve Reise, 2000). Ayrıca MTK'da, KTK'nın aksine bireylerin yetenek düzeyleri için yapılan kestirimlere her bir maddenin katkısı aynı olmayıp yetenek puanları hesaplanırken madde parametrelerindeki farklılıklar da dikkate alınmaktadır (Çelen ve Aybek, 2013). Bununla birlikte; yetenek kestiriminde göz önünde bulundurulmuş parametre sayısı, analiz sırasında MTK kapsamında yer alan bir, iki ve üç parametrelilikten hangisinin işe koşulduğuna bağlı olarak değişiklik göstermektedir. Analizde bir parametrelilik model kullanıldığında yetenek kestiriminde yalnızca madde güçlük

parametresi ( $b$ ) etkili olmaktadır. Analizin iki parametrelili modele göre yapılması halinde, yetenek kestiriminde madde güçlük parametresinin yanı sıra ayırt edicilik parametresi de ( $a$ ) esas alınmaktadır. Analizin üç parametrelili modele göre yürütülmesi durumunda ise yetenek kestirimi sırasında madde güçlük ve ayırt edicilik parametrelerine ek olarak şans parametresi de ( $c$ ) hesaba katılmaktadır (Baker, 2001).

MTK, KTK'nın sınırlılıklarına cevap olabilmesi için geliştirilen bir kuram olduğundan, MTK ileri sürüldükten sonra KTK ile MTK'dan elde edilen sonuçların karşılaştırılmasına yönelik çok sayıda araştırma yapılmıştır. Bu araştırmaların bir kısmında (Awopeju ve Afolabi, 2016; Courville, 2004; Fan, 1998; Güler, Kaya Uyanık ve Taşdelen Teker, 2014; Hwang, 2002; Kan, 2006; Kelecioğlu, 2001; Lawson, 1991; Ojerinde, 2013; Stage, 1998a, 1998b; Uysal, 2015), KTK ile MTK'da rapor edilen madde parametrelerine odaklanılmış ve iki kuramda hesaplanan madde güçlük ile ayırt edicilik değerleri karşılaştırmalı olarak incelemiştir. Çalışmaların bazılarında (Nartgün, 2001; Uysal, 2015), KTK ile MTK'da hesaplanan güvenilirlik katsayılarının ne derece örtüştüğü test edilmiştir. Bunun için KTK'dan elde edilen iç tutarlık katsayısı, MTK'da ölçülen özelliğin farklı düzeyleri için kestirilen güvenilirlik katsayılarının ortalaması olarak kabul edilen marjinal güvenilirlik katsayısıyla karşılaştırılmıştır. KTK ile MTK'nın karşılaştırılmasına yönelik çalışmaların önemli bir kısmında ise (Courville, 2004; Çelen & Aybek, 2013; Fan, 1998; Hwang, 2002; İlhan, 2016; Macdonald ve Paunonen, 2002; Ndalichako ve Todd Rogers, 1997, Özer Özkan, 2014; Zaman, Kashmiri, Mubarak ve Ali, 2008) iki kurama göre elde edilen yetenek kestirimleri arasındaki uyuma bakılmıştır.

KTK ve MTK'da kestirilen yetenek puanlarının karşılaştırılmasına yönelik alanyazındaki çalışmalar incelendiğinde, karşılaştırmaların daha çok başarı testleri üzerinden yapıldığı ve çoğunlukla rapor edilen yetenek puanları arasındaki görece uyumun tespitiyle sınırlı kaldığı anlaşılmaktadır. Ayrıca KTK ve MTK'ya hesaplanan yetenek kestirimlerine ilişkin dağılım özelliklerinin karşılaştırılmasına dönük çalışmaların son derece sınırlı olduğu görülmektedir. Dolayısıyla alanyazında KTK ile MTK'nın karşılaştırıldığı çok sayıda çalışma yer almasına rağmen, iki kuramda hesaplanan yetenek puanlarının ne ölçüde benzer olduğu hakkındaki bir takım soruların henüz tam anlamıyla cevaplanamadığı düşünülmektedir. Öncelikle, alanyazındaki çalışmalarda KTK ve MTK'da hesaplanan yetenek ölçüleri arasındaki uyumun test edilmesinde veri kaynağı olarak genellikle başarı testlerinin kullanılması; bu araştırmalarda ulaşılan, iki kuram arasında yüksek görece uyum olduğu bulgusunu Likert tipi ölçek verilerine genellemeyi güçleştirmektedir. İkinci olarak,

alanyazındaki çalışmalar (İlhan, 2016) açık uçlu sorularla yapılan ölçmelerde MTK'ya dayalı bir model olan çok yüzeyle Rasch modeli ile KTK'da hesaplanan yetenek kestirimleri arasında mutlak bir uyum bulunmadığını ortaya koymakta fakat bu durumun Likert tipi ölçekler için de geçerli olup olmadığına dair bir bilgi vermemektedir. Likert tipi ölçeklerde KTK ve MTK'ya dayalı yetenek kestirimlerinin aynı dağılım özelliklerine sahip olup olmadığı alanyazında yanıtız kalan bir diğere sorudur. Progar ve Sočan (2008) tarafından yapılan çalışmada KTK ve MTK'da hesaplanan yetenek kestirimlerine ait dağılım özellikleri karşılaştırılmıştır. Fakat sözü edilen araştırma başarı testleri üzerinden yürütüldüğünden çalışmada ulaşılan sonuçların Likert tipi ölçek verilerine genellenip genellenemeyeceği bilinmemektedir. Tüm bu gerekçelerle, MTK ve KTK'da hesaplanan yetenek ölçülerinin Likert tipi ölçek verileri üzerinden karşılaştırıldığı bir çalışmanın alanyazına katkı sağlaması beklenmektedir. Bu kapsamda çalışmada, Likert tipi ölçekler için KTK ile bir parametrelili MTK modeli olarak ifade edilen Rasch modelinde hesaplanan yetenek kestirimleri arasındaki uyumun test edilmesi amaçlanmaktadır. Belirtilen bu genel amaç çerçevesinde çalışmada; Likert tipi ölçeklerde KTK ile Rasch modeline göre hesaplanan yetenek kestirimlerinin *i*) görelili uyum, *ii*) mutlak uyum ve *iii*) dağılım özellikleri açısından karşılaştırılması hedeflenmektedir.

### Yöntem

Likert tipi ölçeklerde, klasik test kuramı ile Rasch analizinden elde edilen yetenek kestirimleri arasındaki uyumun test edilmesinin amaçlandığı bu çalışma temel bir araştırma niteliğindedir. Temel araştırmalar, kuramsal araştırmalar olarak da isimlendirilmektedir. Bu araştırmaların öncelikli amacı, yeni bilgilerin üretilmesi ve mevcut bilgi birikiminin genişletilmesidir (Powell, 1997). Kısa süreli bellek ile uzun süreli belleğin kapasitelerinin karşılaştırılması veya beynin herhangi bir bölümündeki doku bozulmasının davranışlar üzerindeki etkisinin test edilmesi temel araştırmalara örnek olarak gösterebilir (Jackson, 2008). Bilginin gerçek yaşam durumlarında kullanılmasına odaklanan uygulamalı çalışmaların aksine, temel araştırmaların uygulamaya yönelme endişesi bulunmamaktadır. Bununla birlikte temel araştırmalarda ulaşılan sonuçlar, dolaylı olarak gerçek yaşam problemlerinin çözümüne katkı sağlayabilmektedir (Martella, Nelson, Morgan ve Marchand Martella, 2013).

### Çalışma Grubu

Araştırma, 200 öğretmenden oluşan bir çalışma grubu üzerinde yürütülmüştür. Katılımcıların 76'sı (%38) kadın ve 124'ü (%62) erkek olup öğretmenlik mesleğindeki hizmet süreleri bir ile 28 yıl arasında değişmektedir. Çalışmanın temel bir araştırma özelliği taşıması, araştırmadaki katılımcı sayısının belirlenmesinde etkili bir rol oynamıştır. Çünkü iç geçerlik (elde edilen

sonuçların doğruluğu), hem uygulamalı hem de temel araştırmalar için son derece önemli iken; dış geçerlik (araştırmada ulaşılan sonuçların genellenebilirliği), uygulamalı araştırmalar için daha fazla önem taşımaktadır (Bickman ve Rog, 2009). Buna bağlı olarak, araştırmadaki katılımcı sayısına karar verilirken dikkate alınan temel ölçüt, KTK'ya ve Rasch modeline ilişkin analizlerin yürütülebileceği büyüklükte bir örnekleme ulaşılabileceği olmuştur. 200 kişilik bir örneklem gerek KTK'ya gerekse de Rasch modeline ilişkin analizler için yeterli kabul edildiğinden (DeMars, 2010; Kline, 1994) çalışma grubundaki katılımcı sayısının daha fazla arttırılmasına gerek görülmemiştir.

### Veri Toplama Aracı

Araştırmanın verileri, Kitsantas (2012) tarafından geliştirip Gezer ve İlhan (2016) tarafından Türkçe'ye uyarlanan *Sınıftaki Farklılıklar için Öğretmen Yeterlikleri Ölçeği (SFÖYÖ)* ile toplanmıştır. Bu ölçek öğretmenlerin, öğretim sürecini sınıftaki farklılıkları göz önünde bulundurabilecek şekilde düzenleyebilme konusundaki yeterlik algılarını ölçmeye yöneliktir. *Çok Güvenirim* (5) ile *Hiç Güvenmem* (1) aralığında uzanan beşli Likert tipi bir derecelendirmeye sahip olan ölçek, 10 madde ve tek faktörden oluşmaktadır. Ölçeğin orijinal ve Türkçe formunda geçerlik ile güvenilirlik analizine ilişkin rapor edilen sonuçlar Tablo 1'de özetlenmiştir.

**Tablo 1.** SFÖYÖ'nün Orijinal ve Türkçe Formunda Geçerlik ile Güvenirlik Analizine İlişkin Rapor Edilen Sonuçlar

	Açımlayıcı Faktör Analizi		Doğrulamalı Faktör Analizi		Güvenirlik
	Açıklanan Varyans	Faktör Yükleri	Uyum İndeksleri	Faktör Yükleri	
Orijinal Form	%61.26	.71 ile .82 arasında	–	–	.91
Türkçe Form	%43.39	.58 ile .77 arasında	$\chi^2/sd=2.13$ , GFI=.93, RFI=.93, CFI=.97, NFI=.94, IFI=.95, NNFI=.96, AGFI=.90, RMSEA=.073, SRMR=.051	.52 ile .74 arasında	.85

SFÖYÖ ile toplanan verilerin bu çalışmadaki güvenilirliği hem KTK'ya hem de Rasch modeline göre hesaplanmıştır. KTK'ya dayalı güvenilirlik kestirimi için Cronbach alfa iç tutarlık katsayısı kullanılmış ve .81 olarak bulunmuştur. Rasch analizinden elde edilen güvenilirlik katsayıları ise birey ve madde yüzeylelerinin her ikisi için de .82 olarak belirlenmiştir.

### Verilerin Analizi

Veri analizi için ilk olarak SFÖYÖ'nün maddelerinden alınan puanlar toplanmış ve ölçekteki madde sayısına bölünerek KTK'ya ilişkin yetenek kestirimleri elde edilmiştir. Ardından birey

ve madde şeklinde iki yüzeyli bir desen ile Rasch analizi gerçekleştirilip bu analizde, logit ölçeğinde rapor edilen yetenek kestirimlerine ulaşılmıştır. Yetenek kestirimlerinin Rasch analizinde logit ölçeğinde rapor edilmesi, KTK'ya ve Rasch modeline göre hesaplanan yetenek kestirimleri arasındaki görel uyumun belirlenmesi için bir engel teşkil etmemektedir. Çünkü görel uyumda, rapor edilen yetenek kestirimleri doğrultusunda katılımcılar arasında bir sıralama yapılmakta ve iki kuramdan elde edilen sıralamaların birbiriyle ne derece tutarlı olduğu test edilmektedir. Öte yandan, KTK ve Rasch analizinde ulaşılan yetenek kestirimleri arasında mutlak bir uyum bulunup bulunmadığı kontrol edilirken katılımcılar arasında yapılan sıralamalar değil; rapor edilen yetenek kestirimlerinin gerçek değerleri esas alınmaktadır. Dolayısıyla KTK ve Rasch analizine ilişkin yetenek kestirimleri arasında mutlak bir uyum olup olmadığının test edilebilmesi için Rasch analizi çıktılarında logit biriminde yer alan yetenek kestirimlerinin SFÖYÖ'de kullanılan dereceleme bilimlerine dönüştürülmesi gerekmektedir. Bu dönüştürme için Linacre'nin (2014) önerdiği formül kullanılmış ve bu formül kapsamında takip edilen adımlar Tablo 2'de özetlenmiştir.

**Tablo 2.** *Logit Ölçeğinde Rapor Edilen Yetenek Kestirimlerinin Likert Ölçeğinde Kullanılan Dereceleme Birimlerine Dönüştürülmesine Yönelik İşlem Adımları*

1. adım	Katılımcıların yetenek düzeylerine ilişkin aralık değeri hesaplanır.
2. adım	Ölçekte benimsenen derecelemeyle ait ranj değeri belirlenir.
3.adım	İkinci adımda bulunan ranj değeri, ilk adımda katılımcıların yetenek düzeyi için hesaplanan aralık değerine bölünür.
4.adım	Üçüncü adımdaki bölme işleminden elde edilen değer ile en düşük yetenek düzeyindeki katılımcıya ait yetenek kestirimi çarpılır.
5.adım	Dördüncü adımda bulunan değer ile toplandığında, ölçekteki derecelemenin en alt ucuna eşit olacak sabit belirlenir.
6.adım	Her bir katılımcı için rapor edilen yetenek kestirimi, ikinci adımda elde edilen değer ile çarpılıp çıkan sonuca beşinci adımdaki sabit eklenir.

Logit ölçeğindeki yetenek kestirimleri Likert tipi ölçekte esas alınan beşli dereceleme birimlerine dönüştürüldükten sonra KTK ve Rasch modeline göre hesaplanan yetenek kestirimlerinin karşılaştırılmasına yönelik istatistiksel işlemlere geçilmiştir. Araştırmada iki kurama göre hesaplanan yetenek kestirimleri arasındaki görel uyumu incelemek amacıyla Pearson korelasyon katsayısından yararlanılmış, mutlak uyumun tespitinde ise ilişkili örneklem *t*-testine başvurulmuştur. Son olarak, gösterdikleri dağılım açısından KTK ve Rasch analizinden elde edilen yetenek kestirimlerinin ne derece uyumlu olduğunu saptamak için histogram grafiği oluşturulmuş, çarpıklık ve basıklık katsayılarına bakılmıştır. Çalışmada; KTK'ya yönelik yetenek kestirimleri, korelasyon analizi, ilişkili örneklem *t*-testi, histogram

grafığı ile çarpıklık ve basıklık katsayıları için SPSS 20.0 paket programından yararlanılmış; Rasch analizinde ise FACETS paket programı kullanılmıştır.

### Bulgular

Bu bölümde, araştırmadan elde edilen bulgulara yer verilmiştir. İlk olarak KTK'ya göre her bir katılımcının ölçekten aldığı puan belirlenmiştir. Katılımcıların yetenek düzeylerine karşılık gelen bu puanlara ait betimleyici istatistikler Tablo 3'te gösterilmiştir. Çalışmada 200 katılımcı bulunduğundan hesaplanan yetenek kestirimleri her bir katılımcı için ayrı ayrı verilememiş; Tablo 3'te yalnızca grubun genelini yansıtan betimleyici istatistikler sunulabilmiştir.

**Tablo 3.** *Katılımcılar için KTK'ya göre Hesaplanan Yetenek Kestirimlerine ait Betimleyici İstatistikler*

n	Ortalama	Standart Sapma	Minimum	Maksimum	Ranj
200	3.51	.63	1.00	5.00	4.00

Tablo 3'e göre, çalışma grubundaki öğretmenlerin öğretim sürecini sınıftaki kültürel farklılıklara hitap edecek şekilde düzenleyebilme konusundaki yeterlik algıları 1.00 ile 5.00 arasında değişmekte ve grup ortalaması 3.51'e karşılık gelmektedir. Grup ortalamasına ilişkin bu değer, sınıftaki kültürel farklılıklara göre şekillenen bir öğretim süreci oluşturabilme konusunda öğretmenlerin kendilerini yeterli gördüklerini yansıtmaktadır. KTK'ya ilişkin yetenek kestirimlerinin hesaplanmasından sonra Rasch analizi uygulanmıştır. Analiz sonucunda birey ve madde yüzeyleri için elde edilen ölçüm raporları Tablo 4'te sunulmuştur. KTK'da olduğu gibi Rasch analizinde de hesaplanan yetenek kestirimleri çalışma grubundaki her bir katılımcı için ayrı ayrı sunulamamış; yalnızca grubun geneline ilişkin ortalama değer şeklinde verilmiştir.

**Tablo 4.** *Rasch Analizinde Birey ve Madde Yüzeyi için Elde Edilen Ölçüm Raporları*

	Birey Yüzeyi			Madde Yüzeyi		
	Logit ölçüsü	Uyum-içi	Uyum-dışı	Logit ölçüsü	Uyum-içi	Uyum-dışı
Ortalama	.56	1.02	1.02	.00	1.00	1.02
Standart Sapma	1.02	.75	.75	.21	.15	.18
Ayırma Oranı		2.10			2.16	
Güvenirlilik		.82			.82	
Serbestlik Derecesi		199			9	
Ki Kare		851.10			49.4	

Tablo 4'e bakıldığında, öğrenme ortamını sınıftaki kültürel farklılıklara uygun şekilde düzenleyebilme konusundaki yeterlik algıları açısından öğretmenlerin anlamlı düzeyde



birbirinden ayırt edildiği görülmektedir [ $\chi^2_{(sd=199)} = 851.10, p < .01$ ]. SFÖYÖ'deki maddelerin güçlük düzeyleri arasında anlamlı fark bulunduğu Tablo 4'ün ortaya koyduğu bir diğer önemli bulgudur [ $\chi^2_{(sd=9)} = 49.4, p < .01$ ]. Buna göre, SFÖYÖ'deki bazı maddelerde *çok güvenirim* ya da *güvenirim* kategorilerinin işaretlenme olasılığı daha yüksek iken, bazı maddelerde *hiç güvenmem* ya da *güvenmem* kategorilerinin işaretlenme olasılığının daha yüksek olduğu söylenebilir. Tablo 4'teki uyum istatistikleri incelendiğinde hem uyum-içi hem de uyum-dışı istatistiklerinin 1.00'a oldukça yakın olduğu belirlenmiştir. Uyum istatistiklerinin 0.6 ile 1.4 aralığında bulunması, model ile verinin uyumlu olduğuna işaret etmektedir (Wright ve Linacre, 1994). Dolayısıyla model-veri uyumunun sağlandığı söylenebilir. Rasch analizinin ardından, logit birimde rapor edilen yetenek kestirimlerini SFÖYÖ'de kullanılan beşli derecelemenin birimlerine dönüştürmeye yönelik işlemlere geçilmiş ve bu işlemler Tablo 5'te özetlenmiştir.

**Tablo 5.** Rasch Analizinde Logit Biriminde Rapor Edilen Yetenek Kestirimlerinin SFÖYÖ'de Kullanılan Beşli Derecelemenin Birimlerine Dönüştürülmesi

1. adım	Öğrenme sürecini sınıftaki kültürel farklılıklara hitap edecek şekilde düzenleyebilme konusunda en yüksek ve en düşük yeterlik algısına sahip öğretmenler için hesaplanan yetenek kestirimleri sırasıyla 5.66 ile -5.05 logit'tir. Dolayısıyla katılımcıların yetenek düzeylerine ilişkin aralık 10.71'dir.
2. adım	SFÖYÖ'de <i>Çok Güvenirim</i> (5) ile <i>Hiç Güvenmem</i> (1) şeklinde beşli bir dereceleme kullanılmıştır. Bu sebeple, ölçekteki derecelemeyle ilişkin ranj değeri $5-1 = 4$ 'tür.
3. adım	Ölçekteki derecelemeyle ilişkin ranj değeri, katılımcıların yetenek düzeyine ilişkin aralığa bölüldüğünde 0.373 değeri elde edilmiştir [ $4 \div 10.71 = 0.373$ ].
4. adım	Üçüncü adımdaki bölme işleminden elde edilen 0.373 değeri, ölçülen özellik bakımından en düşük yetenek düzeyine sahip olan öğretmenin yetenek kestirimi ile çarpıldığında $-1.886$ değerine ulaşılmıştır [ $(-5.05) \times (0.373) = -1.886$ ].
5. adım	Dördüncü adımda bulunan değer ile toplandığında SFÖYÖ'deki derecelemenin alt ucuna (1) karşılık gelecek sabiti belirlemek için $(-1.886) + (\text{sabit}) = 1$ eşitliği oluşturulmuştur. Bu eşitliğe göre, sabit değeri 2.886 olarak elde edilmiştir.
6. adım	Buna göre, logit birimindeki yetenek kestirimlerinin SFÖYÖ'deki beşli derecelemenin birimlerine dönüştürülmesinde kullanılacak formül aşağıdaki gibidir. $\left( \text{Katılımcıların Logit} \right) \times (0.373) + 2.886$ $\left( \text{Ölçeğindeki Yetenej Düzeyi} \right)$

Tablo 5'in son satırında ulaşılan formül kullanılarak Rasch analizi çıktılarında logit biriminde sunulan yetenek kestirimleri SFÖYÖ'de benimsenen beşli derecelemenin birimlerine dönüştürülmüştür. Bu dönüştürme işlemiyle birlikte, KTK ile Rasch analizinden elde edilen yetenek kestirimleri karşılaştırmaya hazır hale gelmiştir. İki kurama göre hesaplanan yetenek kestirimleri arasındaki görel uyumu belirlemek için uygulanan korelasyon analizi ile mutlak

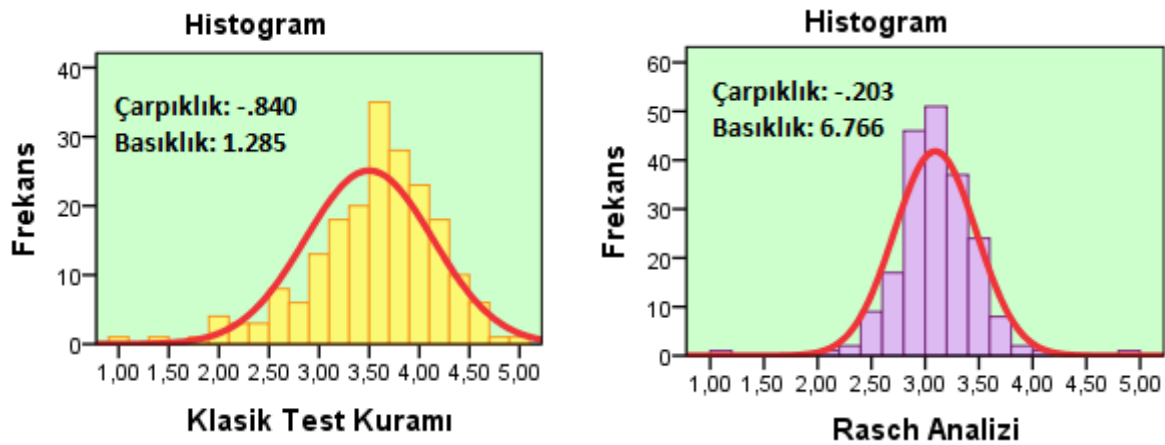
uyumu test etmek amacıyla gerçekleştirilen ilişkili örneklem *t*-testi sonuçları Tablo 6’da bir arada sunulmuştur.

**Tablo 6.** *KTK İle Rasch Analizinden Elde Edilen Yetenek Kestirimleri Arasındaki Uyumun Belirlenmesine Yönelik Korelasyon Analizi ile İlişkili Örneklem t-Testi Sonuçları*

Kuram	Ortalama	Standart Sapma	N	<i>r</i>	<i>t</i>
KTK	3.50	.38	200	.95	19.30
Rasch	3.09	.64			

Tablo 6’da görüldüğü üzere, KTK ile Rasch analizinden elde edilen yetenek kestirimleri arasında pozitif yönde ve oldukça güçlü bir ilişki bulunmaktadır [ $r = .95, p < .01$ ]. Korelasyon analizine ilişkin bu sonuç, iki kurama göre hesaplanan yetenek kestirimleri arasında yüksek bir görelî uyum bulunduğunu yansıtmaktadır. İlişkili örneklem *t*-testi sonuçlarına bakıldığında ise KTK’dan elde edilen yetenek kestirimlerinin Rasch analizinde rapor edilen yetenek kestirimlerine kıyasla anlamlı derecede daha yüksek olduğu anlaşılmaktadır [ $t_{(199)} = 19.30, p < .01$ ]. Buna göre, iki kuramda ulaşılan yetenek kestirimleri arasında mutlak bir uyumdan söz edilmesi mümkün değildir.

KTK ile Rasch analizinden elde edilen yetenek kestirimleri arasındaki görelî ve mutlak uyum inceledikten sonra, gösterdikleri dağılım açısından iki kurama göre hesaplanan yetenek kestirimlerinin ne derece örtüştüğü kontrol edilmiştir. Bu kapsamda oluşturulan histogram grafikleri, çarpıklık ve basıklık katsayıları ile birlikte Şekil 1’de gösterilmiştir.



**Şekil 1.** *KTK ile Rasch Analizinden Elde Edilen Yetenek Kestirimlerine ait Histogram Grafikleri, Çarpıklık ve Basıklık Katsayıları*

Şekil 1'deki histogram grafikleri hem KTK'dan hem de Rasch analizinden elde edilen yetenek kestirimlerinin normalden önemli bir sapma göstermediğine işaret etmektedir. Bununla birlikte, çarpıklık katsayısının KTK'da  $-1$ 'e; Rasch analizinde ise  $0$ 'a daha yakın olması, Rasch analizinde rapor edilen yetenek kestirimlerinin KTK'dan hesaplanan yetenek kestirimleriyle kıyasla normale daha yakın olduğunu göstermektedir. Şekil 1'e göre ayrıca KTK'ya ve Rasch analizine göre hesaplanan yetenek kestirimlerinin basıklıkları arasında da farklılık bulunmaktadır. Hesaplanan basıklık katsayıları her iki kuram için de sıfırdan büyük olmasına rağmen, Rasch analizinde rapor edilen yetenek kestirimleri için basıklık katsayısı daha yüksek bulunmuştur. Basıklık katsayıları arasındaki bu fark, KTK'ya göre elde edilen yetenek kestirimlerinin Rasch analizinde rapor edilen yetenek kestirimlerine göre daha heterojen olduğu anlamına gelmektedir.

### **Tartışma, Sonuç ve Öneriler**

Bu araştırmada, Likert tipi ölçeklerde KTK ile Rasch analizinden elde edilen yetenek kestirimleri arasındaki uyumun test edilmesi amaçlanmıştır. Bu amaç doğrultusunda öncelikle, iki kurama göre hesaplanan yetenek kestirimleri arasındaki görelî uyum incelenmiştir. Korelasyon analizi sonucunda KTK ile Rasch analizinde rapor edilen yetenek kestirimleri arasında pozitif yönde ve oldukça güçlü bir ilişki bulunduğu belirlenmiştir. Buna göre, Likert tipi ölçeklerde KTK ile Rasch analizinden elde edilen yetenek kestirimleri arasındaki görelî uyumun son derece yüksek olduğu sonucuna varılmıştır. Bu sonuç; Likert tipi ölçeklerle toplanan veriler ister KTK'ya ister Rasch modeline göre analiz edilsin ulaşılan yetenek kestirimleri açısından katılımcılar arasında yapılan sıralamalarda bir farklılık olmadığı anlamına gelmektedir. Araştırmaya ilişkin bu bulgu, MTK ile KTK'dan elde edilen yetenek kestirilmesinin karşılaştırılmasına yönelik farklı çalışmalarda ulaşılan sonuçlar ile aynı doğrultudadır. Örneğin; Ndalichako ve Todd Rogers (1997) tarafından yapılan araştırmada 70 maddelik çoktan seçmeli bir teste 1232 öğrencinin verdiği cevaplar KTK ile bir, iki ve üç parametrelî MTK modellerine göre analiz edilmiştir. Çalışmada KTK'dan elde edilen yetenek kestirimlerinin bir, iki ve üç parametrelî MTK modellerinde hesaplanan yetenek kestirimleri ile arasındaki korelasyon katsayıları sırasıyla .977, .985 ve .988 olarak bulunmuştur. Fan'ın (1998) okuma ve matematik testlerini kullanılarak yaptığı çalışmada; KTK ile bir, iki ve üç parametrelî MTK'da rapor edilen yetenek kestirimleri arasındaki ortalama korelasyon değeri .96 olarak tespit edilmiştir. Macdonald ve Paunonen (2002) tarafından yapılan çalışmada ise KTK ve MTK'da hesaplanan yetenek kestirimleri, Monte

Carlo yöntemi kullanılarak üretilen simülasyon verileri üzerinden karşılaştırılmıştır. Araştırmada, 1000 kişilik bir örneklem için farklı test uzunluğuna, farklı madde güçlük ve ayırt edicilik indekslerine sahip çeşitli deneysel koşullar oluşturulmuş ve tüm deneysel koşullarda KTK ile MTK'ya göre kestirilen yetenek parametreleri arasındaki korelasyon katsayılarının .970 ile .995 arasında değiştiği saptanmıştır. Çelen ve Aybek (2013) tarafından yapılan çalışmada, sınıf içi ders başarısını ölçmeyi amaçlayan bir ölçme aracından KTK ile MTK'ya dayalı olarak hesaplanan öğrenci başarıları arasındaki ilişki test edilmiş ve iki kurama göre elde edilen yetenek kestirimleri arasındaki korelasyon katsayısı .95 şeklinde bulunmuştur. Özer Özkan (2014) tarafından yapılan çalışmada da benzer bulgulara ulaşılmış olup öğrenci başarısının belirlenmesi sınavından KTK ile tek ve çok boyutlu MTK'ya göre kestirilen başarı puanları arasında pozitif yönde güçlü ilişkilerin bulunduğunu saptamıştır. Tüm bu sonuçlar KTK ile MTK'ya göre hesaplanan yetenek ölçülerinin bireyleri benzer şekilde sıraya koyduğuna işaret etmekte ve bu araştırmada ulaşılan bulguları desteklemektedir.

Likert tipi ölçeklerde KTK ve Rasch modeline göre hesaplanan yetenek kestirimleri arasındaki yüksek görece uyumun aksine, ilişkili örneklem *t*-testi sonuçları iki kurama göre kestirilen yetenek puanları arasında mutlak bir uyumdan söz etmenin mümkün olmadığını ortaya koymuştur. KTK'dan elde edilen yetenek kestirimleri Rasch analizinde rapor edilen yetenek kestirimlerine kıyasla anlama derecede daha yüksek çıkmıştır. Bu bulgu, İlhan (2016) tarafından yapılan çalışmanın sonuçları ile kısmen paralellik göstermektedir. İlhan'ın (2016) yaptığı çalışmada açık uçlu sorularla yapılan ölçmelerde KTK ile çok yüzeyli Rasch analizinden elde edilen yetenek kestirimleri karşılaştırılmıştır. Çalışmada, 100 öğrencinin açık uçlu altı soruya verdiği cevapların dört puanlayıcı tarafından puanlanmasıyla elde edilen veriler KTK ile çok yüzeyli Rasch modeline göre analiz edilmiştir. Analiz çıktıları karşılaştırıldığında çok yüzeyli Rasch modelinde hesaplanan yetenek kestirimlerinin KTK'da hesaplanan yetenek kestirimlerine kıyasla anlamlı derecede daha yüksek olduğu tespit edilmiştir. Bu bağlamda, KTK ve Rasch modeline göre hesaplanan yetenek kestirimleri arasında mutlak bir uyum bulunmadığını ortaya koyması yönüyle bu araştırmada ulaşılan bulgular ile İlhan (2016) tarafından yapılan çalışmada ulaşılan sonuçlar aynı eksendedir. Ancak iki kurama göre elde edilen yetenek kestirimleri arasında saptanan anlamlı farkın, bu çalışmada KTK'nın lehine iken; İlhan'ın (2016) yaptığı çalışmada çok yüzeyli Rasch modelinin lehine olması iki araştırmada ulaşılan sonuçları birbirinden farklı kılmaktadır. Söz konusu fark, iki kuramın bu çalışmada Likert tipi ölçek verileri üzerinden karşılaştırılırken,

İlhan'ın (2016) yaptığı araştırmada açık uçlu sorular üzerinden karşılaştırılmasından kaynaklanıyor olabilir. Çünkü Likert tipi ölçekler Rasch modeline göre analiz edildiğinde ölçme sonuçları üzerinde etkili olan değişkenlik kaynakları bireyler ve maddeler ile sınırlı kalmakta ve analizler iki yüzeyle bir desen kullanılarak gerçekleştirilmektedir. Oysa açık uçlu sorularda, bireyler ve maddelerin yanı sıra puanlayıcılar da ölçme sonuçlarında farklılaşmaya yol açabilen bir değişkenlik kaynağı olarak işlem görmekte ve Rasch analizi bireyler, maddeler ve puanlayıcılar şeklinde üç yüzeyle bir desen ile yürütülmektedir. Analizde işlem gören yüzey sayılarının farklı olması bu araştırmada ulaşılan bulgular ile İlhan'ın (2016) çalışmasında ulaşılan sonuçlar arasındaki farklılığa sebep olan faktörlerden biri olabilir. Bununla birlikte; farklı ölçme kuramlarının karşılaştırılmasına yönelik çalışmalarda veri setlerinden kaynaklı etkiler de mümkün olabildiğinden (Engelhard, 1984), bu konuda daha kesin bir yargıya varılabilmesi için hem Likert tipi ölçeklerde hem de açık uçlu sorularda KTK ve Rasch modelinin karşılaştırılmasına yönelik çalışma sayısının arttırılmasına ihtiyaç olduğu düşünülmektedir.

Araştırmada ulaşılan üçüncü bir sonuç, Likert tipi ölçeklerde KTK ve Rasch modeline göre hesaplanan yetenek kestirimlerinin dağılım özellikleri ile ilgilidir. Elde edilen sonuçlara göre, KTK ile Rasch analizinden elde edilen çarpıklık katsayılarının her ikisi de negatiftir ve verilerin normalden önemli bir sapma göstermediğini söyleyebilmek için dikkate alınması önerilen  $\pm 1$  aralığının (Büyüköztürk, 2010) içinde kalmaktadır. Fakat çarpıklık katsayılarının KTK'ya göre yapılan analizde -1'e; Rasch modeline göre gerçekleştirilen analizde ise 0'a daha yakın olduğu da gözden kaçırılmamalıdır. Dolayısıyla, gösterdikleri dağılım açısından iki kurama göre hesaplanan yetenek kestirimleri arasında bir benzerlik bulunsa da tam bir örtüşmenin olmadığı ifade edilebilir. Bu durum basıklık katsayıları için de geçerlidir. KTK ve Rasch modeline göre hesaplanan yetenek kestirimlerine ait basıklık katsayılarının her ikisi de 0'dan yüksek olmakla birlikte, Rasch modelinde rapor edilen yetenek kestirimlerine ilişkin basıklık katsayısı KTK'da ulaşılan yetenek kestirimlerine ilişkin basıklık katsayısından önemli ölçüde daha yüksek bulunmuştur. Bu noktadan hareketle, KTK'da rapor edilen yetenek kestirimleri arasındaki değişkenliğin Rasch modeline göre hesaplanan yetenek kestirimleri arasındaki değişkenlikten daha fazla olduğu söylenebilir. KTK ve Rasch modelinde hesaplanan yetenek kestirimlerinin dağılım özelliklerine ilişkin ulaşılan sonuçlar, Progar ve Sočan (2008) tarafından yapılan çalışmanın sonuçlarıyla benzer niteliktedir. Progar ve Sočan (2008) Üçüncü Uluslararası Fen ve Matematik Eğitimi Araştırmasına [Third International Mathematics and Science Study (TIMSS 1995)] ait veriler üzerinden KTK ve

MTK analizlerine ilişkin sonuçları karşılaştırmıştır. Çalışmada KTK ve MTK'ya göre kestirilen yetenek puanlarının her ikisi de sola (negatif) çarpık olmasına karşın MTK'da rapor edilen yetenek kestirimlerinin KTK'da hesaplanan yetenek kestirimlerine kıyasla daha simetrik olduğu belirlenmiştir. Ayrıca, MTK'da ulaşılan yetenek kestirimlerinin KTK'dan elde edilen yetenek kestirimlerine göre daha sivri bir yapıya sahip olduğu saptanmıştır.

Dağılım özellikleri açısından KTK ve Rasch modeline göre hesaplanan yetenek kestirimleri arasında gözlenen farklılıklar; sıralama düzeyinde olan Likert tipi ölçek verilerine, KTK ve Rasch analizinde değişik işlemler uygulanmasıyla açıklanabilir. Likert tipi ölçeklerden elde edilen orijinal veriler sıralama ölçeğinde olmasına ve normal şartlarda toplama/çıkarma gibi matematiksel işlemler yapmaya uygun olmamasına rağmen, KTK'da bu veriler eşit aralıklı ölçek düzeyindeymiş gibi işlem görmekte ve bireylerin yetenek kestirimlerini elde etmek için ölçeğin her bir maddesinden aldığı puanlar toplanmaktadır. Rasch modelinde ise Likert tipi ölçekler eşit aralıklı logit ölçeğine dönüştürülmekte ve katılımcıların yetenek düzeyleri logit biriminde rapor edilmektedir. KTK ve Rasch modelinin Likert tipi ölçek verilerine yaklaşımlarındaki bu farklılık, iki kurama hesaplanan yetenek kestirimlerine ait çarpıklık ve basıklık katsayılarının tam olarak örtüşmemesine yol açmış olabilir. Nitekim Jamieson (2004) ile Brinthaup ve Kang (2014), Likert tipi ölçek verileri sıralama düzeyinde olduğundan bu ölçeklerden alınan toplam puanların yanlı sonuçlar üretebileceğini belirtmiştir. KTK'ya göre bireyin yetenek kestirimine tüm maddeler eşit katkı sağlarken; MTK'ya dayalı modellerde yetenek ölçülerinin madde parametreleri göz önünde bulundurularak hesaplanması iki kuram için elde edilen çarpıklık ve basıklık katsayılarındaki farklılığa kaynaklık eden bir başka neden olabilir. Yine MTK'da bireylerin yetenek düzeyleri madde parametrelerinden bağımsız olarak hesaplanırken; KTK'da hesaplanan yetenek düzeylerinin madde parametrelerine bağımlı olması, iki kuramdan elde edilen yetenek ölçülerinin dağılım özelliklerinin farklılık göstermesine yol açmış olabilir. Daha açık bir anlatımla, KTK'da bireylerin yetenek düzeyleri madde parametrelerine bağımlı olduğundan zor maddelerden oluşan bir ölçme aracında sağa çarpık, kolay maddelerden oluşan bir ölçme aracında ise sola çarpık bir dağılım elde edilme olasılığı daha yüksektir. MTK'da ise bireylerin yetenek düzeyleri madde parametrelerinden bağımsız olarak kestirilebildiğinden, hesaplanan yetenek kestirimlerinin dağılım özelliklerinde maddelerin güçlük düzeylerine bağlı farklılıklar oluşması daha zayıf bir ihtimal haline gelebilmektedir.

Özetle, Likert tipi ölçekler için KTK ve Rasch analizinde rapor edilen yetenek kestirimleri bireyleri neredeyse tamamıyla aynı şekilde sıraya koymakta ama iki kuramdan

hesaplanan yetenek ölçülerinin gerçek değerleri istatistiksel açıdan anlamlı farklılık göstermektedir. Ayrıca iki kuramda ulaşılan yetenek kestirimleri arasında, dağılımın yönü açısından bir tutarlılık bulunsa da hesaplanan çarpıklık ve basıklık katsayıları arasında dikkate değer bir farklılık söz konusudur. Bu sonuçlar, alanyazındaki mevcut çalışmalar ile birlikte değerlendirildiğinde farklı ölçme kuramlarına göre yürütülen analizlerin bireyler hakkında alınacak bağıl kararlarda bir farka sebep olmayacağı şeklinde genel bir değerlendirme yapmak mümkündür. Ancak böylesi genel bir değerlendirmenin KTK ile Rasch analizinde rapor edilen yetenek kestirimleri arasındaki mutlak uyum ve dağılım özellikleri için yapılması pek olanaklı görünmemektedir. Alanyazında KTK ile MTK'da hesaplanan yetenek kestirimlerine ait dağılım özelliklerinin karşılaştırıldığı ve bu iki ölçme kuramında rapor edilen yetenek kestirimleri arasındaki mutlak uyumun test edildiği çalışmaların (Progar ve Sočan, 2008; İlhan, 2016) sayıca az olması, bu tür genel bir değerlendirme yapılmasını güçleştirmektedir. Bu bakımdan, KTK ile Rasch analizinde rapor edilen yetenek kestirimleri arasındaki mutlak uyumun sınanmasına ve iki kuramda hesaplanan çarpıklık ve basıklık katsayılarının karşılaştırılmasına yönelik yeni çalışmaların yapılması, araştırmada ulaşılan bulguların genellenebilirliğine katkı sağlayacaktır.

### Kaynakça

- Anshel, M.H., Weatherby, N.L., Kang, M. ve Watson, T. (2009). Rasch calibration of a unidimensional perfectionism inventory for sport. *Psychology of Sport and Exercise*, 10(1), 210-216. <http://dx.doi.org/10.1016/j.psychsport.2008.07.006>
- Awopeju, O.A. ve Afolabi, E.R.I. (2016). Comparative analysis of classical test theory and item response theory based item parameter estimates of senior school certificate mathematics examination. *European Scientific Journal*, 12(28), 263-284. <http://dx.doi.org/10.19044/esj.2016.v12n28p263>
- Baker, F.B. (2001). *The basics of item response theory*. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD. [http://hbanaszak.mjr.uw.edu.pl/TempTxt/Baker\\_2001\\_TheBasicsOfItemResponseTheory.pdf](http://hbanaszak.mjr.uw.edu.pl/TempTxt/Baker_2001_TheBasicsOfItemResponseTheory.pdf) adresinden elde edildi.
- Bickman, L. ve Rog, D.R. (2009). *Handbook of applied social research methods*. Thousand Oaks, CA: Sage Publications.
- Brinthaup, T.M. ve Kang, M. (2014). Many-faceted rasch calibration: An example using the self-talk scale. *Assessment*, 21(2) 241-249. <http://dx.doi.org/10.1177/1073191112446653>
- Büyüköztürk, Ş. (2010). *Sosyal bilimler için veri analizi el kitabı*. Ankara: Pegem Akademi Yayıncılık.
- Courville, T.G. (2004). *An empirical comparison of item response theory and classical test theory item/person statistics* (Unpublished doctoral dissertation). Texas A&M University, Texas. <http://oaktrust.library.tamu.edu/bitstream/handle/1969.1/1064/etd-tamu-2004B-EPSY-Courville-2.pdf?sequence=1&isAllowed=y> adresinden elde edildi.
- Çelen, Ü. ve Aybek, E.C. (2013). Öğrenci başarısının öğretmen yapımı bir testle klasik test kuramı ve madde tepki kuramı yöntemleriyle elde edilen puanlara göre karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 4(2), 64-75. <http://dergipark.ulakbim.gov.tr/epod/article/view/1040000004> adresinden elde edildi.
- Çıkrıkçı Demirtaşlı, N. (1996). Test geliştirmede yeni yaklaşımlar: Örtük özellikler kuramı-Temel özellikleri, varsayımları, modeller ve sınırlılıklar. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 28(2), 434-445. <http://dergiler.ankara.edu.tr/dergiler/40/486/5705.pdf> adresinden elde edildi.
- DeMars, C. (2010). *Item response theory*. Oxford, UK: Oxford University Press.



Doğan, N. ve Tezbaşaran, A. (2003). Klasik test kuramı ve örtük özellikler kuramının örneklem bağlamında karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 25, 58-67.

<http://dergipark.ulakbim.gov.tr/hunefd/article/download/5000048790/5000046110>

adresinden elde edildi.

Embretson, S.E. ve Reise, S.P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Engelhard, G. (1984). Thorndike, Thurstone, and Rasch: A comparison of their methods of scaling psychological and educational tests. *Applied Psychological Measurement*, 8(1), 21-38. <http://dx.doi.org/10.1177/014662168400800104>

Fan, X. (1998). Item response theory and classical test theory: An empirical comparison of their item/person statistics. *Educational and Psychological Measurement*, 58(3), <http://dx.doi.org/10.1177/0013164498058003001>

Gezer, M. ve İlhan, M. (2016). Sınıftaki farklılıklar için öğretmen yeterlikleri ölçeğinin Türkçeye uyarlaması. *Hasan Ali Yücel Eğitim Fakültesi Dergisi*, 13(2), 187-200. <http://www.journals.istanbul.edu.tr/iuayefd/article/viewFile/5000146180/5000170741>

adresinden elde edildi.

Güler, N., Kaya Uyanık G. ve Taşdelen Teker, G. (2014). Comparison of classical test theory and item response theory in terms of item parameters. *European Journal of Research on Education*, 2(1), 1-6. <http://iassr2.org/rs/020101.pdf> adresinden elde edildi.

Haiyang, S. (2010). An application of classical test theory and many facet Rasch measurement in analyzing the reliability of an English test for non-English major graduates. *Chinese Journal of Applied Linguistics*, 33(2), 87-102. <http://www.celea.org.cn/teic/90/10060807.pdf> adresinden elde edildi.

Hambleton, R.K. ve Jones, R.W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12(3), 38-47. <http://dx.doi.org/10.1111/j.1745-3992.1993.tb00543.x>

Hambleton, R.K., Swaminathan, H. ve Rogers, H.J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: SAGE Publications, Inc.

Hwang, (2002, February). *Classical test theory and item response theory: Analytical and empirical comparisons*. Paper presented at the Annual Meeting of the Southwest

Educational Research Association, Austin, TX.

<http://files.eric.ed.gov/fulltext/ED466779.pdf> adresinden elde edildi.

İlhan, M. (2016). Açık uçlu sorularla yapılan ölçmelerde klasik test kuramı ve çok yüzeyli Rasch modeline göre hesaplanan yetenek kestirimlerinin karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 31(2), 346-368. <http://dx.doi.org/10.16986/HUJE.2016015182>

Jackson, S. (2008). *Research methods: A modular approach*. Belmont, CA: Thompson Wadsworth.

Jamieson, S. (2004). Likert scales: How to (ab)use them. *Medical Education*, 38(12), 1212-1218. <http://dx.doi.org/10.1111/j.1365-2929.2004.02012.x>

Kan, A. (2006). Klasik test teorisine ve örtük özellikler teorisine göre kestirilen madde parametrelerinin karşılaştırılması üzerine ampirik bir çalışma. *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, 2(2), 227-235. <http://dergipark.ulakbim.gov.tr/mersinefd/article/view/5000002951/0> adresinden elde edildi.

Kelecioğlu, H. (2011). Örtük özellikler teorisindeki b ve a parametreleri ile klasik test teorisindeki p ve r istatistikleri arasındaki ilişki. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 20, 104-110. <http://www.efdergi.hacettepe.edu.tr/yonetim/icerik/makaleler/992-published.pdf> adresinden elde edildi.

Kitsantas, A. (2012). Teacher efficacy scale for classroom diversity (TESCD): A validation study. *Revista de Currículum Formación de Profesorado*, 16(1), 35-44. <http://www.ugr.es/~recfpro/rev161ART3en.pdf> adresinden elde edildi.

Kline, R.B. (1994). *An easy guide to factor analysis*. New York: Routledge.

Lawson, S. (1991). One parameter latent trait measurement: Do the results justify the effort? In B. Thompson (Ed.), *Advances in educational research: Substantive findings, methodological developments* (Vol. 1, pp. 159-168). Greenwich, CT: JAI Press.

Linacre, J.M. (2014). *A user's guide to FACETS Rasch-model computer programs*. <http://www.winsteps.com/a/facets-manual.pdf> adresinden elde edildi.

Macdonald, P. ve Paunonen, S.V. (2002). A Monte Carlo comparison of item and person statistics based on item response theory versus classical test theory. *Educational and*

Martella, R.C., Nelson, J.R., Morgan, R.L. ve Marchand Martella, N.E. (2013).

*Understanding and interpreting educational research*. New York: The Guilford Press.

Nartgün, Z. (2001). Klasik test kuramındaki Cronbach alfa güvenilirlik katsayısı ile örtük özellikler kuramındaki bilgi fonksiyonları ve marjinal güvenilirlik katsayısının karşılaştırılması. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 1(2), 143-161. <http://dergipark.ulakbim.gov.tr/aibuefd/article/view/5000091592> adresinden elde edildi.

Ndalichako, J.L. ve Todd Rogers, W. (1997). Comparison of finite state score theory, classical test theory, and item response theory in scoring multiple choice items.

*Educational and Psychological Measurement*, 57(4), 580-589.

<http://dx.doi.org/10.1177/0013164497057004004>

Ojerinde, D. (2013, May). *Classical test theory (CTT) vs item response theory (IRT): An evaluation of comparability of item analysis results*. Lecture Presentation at the Institute of Education, University of Ibadan, Nijerya.

[http://ui.edu.ng/sites/default/files/PROF%20OJERINDE'S%20LECTURE%20\(Autosaved\).pdf](http://ui.edu.ng/sites/default/files/PROF%20OJERINDE'S%20LECTURE%20(Autosaved).pdf) adresinden elde edildi.

Özer Özkan, Y. (2014). Öğrenci başarılarının belirlenmesi sınavından klasik test kuramı, tek ve çok boyutlu madde tepki kuramı modelleri ile kestirilen başarı puanlarının karşılaştırılması. *International Journal of Human Sciences*, 11(1), 20-44.

<http://dx.doi.org/10.14687/ijhs.v11i1.2739>

Powell, R.R. (1997). *Basic research methods for librarians*. Greenwich, CT: Ablex Publishing.

Progar, S. ve Sočan, G. (2008). An empirical comparison of item response theory and classical test theory. *Psihološka obzorja / Horizons of Psychology*, 17(3), 5-24.

<http://www.dlib.si/stream/URN:NBN:SI:DOC-0QF7AGYU/f6e1e8cf-6c06-4644-bf7f-922938b1dc74/PDF> adresinden elde edildi.

Stage, C. (1998a). *A comparison between item analysis based on item response theory and classical test theory. A study of the SweSAT subtest WORD*. (Educational Measurement No 29). Umeå University, Department of Educational Measurement.

[http://www.edusci.umu.se/digitalAssets/60/60607\\_enr2998sec.pdf](http://www.edusci.umu.se/digitalAssets/60/60607_enr2998sec.pdf) adresinden elde edildi.

Stage, C. (1998b). *A comparison between item analysis based on item response theory and classical test theory. A study of the SweSAT subtest ERC*. (Educational Measurement No 30). Umeå University, Department of Educational Measurement. [http://www.edusci.umu.se/digitalAssets/60/60608\\_enr3098sec.pdf](http://www.edusci.umu.se/digitalAssets/60/60608_enr3098sec.pdf) adresinden elde edildi.

Uysal, M. (2015). *Araştırma özyeterlik ölçeğinin psikometrik özelliklerinin klasik test kuramı ve madde tepki kuramına göre incelenmesi* (Yayınlanmamış yüksek lisans tezi). Gazi Üniversitesi, Ankara.

Van der Linden, W.J. ve Hambleton, R.K. (1997). *Handbook of modern item response theory*. New York: Springer.

Wright, B.D. ve Linacre, J.M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8, 370-371. <http://www.rasch.org/rmt/rmt83b.htm> adresinden elde edildi.

Zaman, A., Kashmiri, A., Mubarak, M. ve Ali, A. (2008, November). *Students ranking, based on their abilities on objective type test: Comparison of CTT and IRT*. EDU-COM International Conference, Edith Cowan University, Perth Western Australia <http://ro.ecu.edu.au/ceducom/52/> adresinden elde edildi.

### Extended Abstract

On examining the historical development of the science of measurement and evaluation, it becomes apparent that basically two test theories are available- namely, Classical Test Theory (CTT) and Item Response Theory (IRT). IRT is a theory which was developed in response to the restrictions of CTT, and a great number of studies comparing the ability scores estimated in CTT and in IRT were conducted after IRT was suggested. A review of those studies made it clear that the comparisons were made rather on the basis of achievement tests and that they were mostly restricted to identifying relative agreements between reported ability scores. Besides, it is thought that the number of studies comparing the distribution properties of ability estimations calculated in CTT and in IRT is quite limited. Therefore, despite the availability of a great number of studies comparing CTT and IRT in the literature, it is believed that some of the questions about the extent to which the ability scores calculated in the two theories are similar have not been fully answered yet. Firstly, using achievement tests as a source of data in testing the agreement between ability measures in CTT and in IRT calculated in the literature makes it difficult to generalize those studies' finding that there is a high level of relative agreement between the two theories into the data of Likert type scales. Secondly, studies in the literature (İlhan, 2016) demonstrate that there are no absolute agreements between ability estimations in many facet Rasch Model – which is a model based on IRT- and the ones in CTT in measurements made with open-ended questions, but no information is available in those studies on whether or not the situation is true for Likert type studies. Another question which is unanswered in relevant literature is whether or not ability estimations based on CTT and on IRT have the same distribution. Progar and Sočan (2008) compare the distribution properties for ability estimations calculated in CTT with those calculated in IRT. Yet, whether the results obtained will be generalized into Likert type scale data or not is unknown since the study was conducted with achievement tests. For all these reasons, it is expected that a study in which ability measures calculated in IRT and in CTT are compared on the basis of Likert type scale data would contribute to the literature Hence, this study aimed to compare the ability estimations calculated in CTT with the ones calculated in Rasch model-which is a one-parameter IRT model- in terms of *i*) relative agreement, *ii*) absolute agreement, and *iii*) distribution properties.

Research data were collected from a study group of 200 teachers by means of Teacher Efficacy Scale for Classroom Diversity (TESCD) containing 10 items with one-factor structure. First, the scores received from the items of TESCD were added up and then the total

was divided into the number of items in the scale, and thus ability estimations for CTT were made. After that, Rasch analysis was performed and ability estimations reported in the logit scale were made. After the ability estimations reported in the logit unit were transformed into the units of five-pointed grading used in TESCD, the ability estimations calculated in these two theories were ready for comparison. Pearson's correlations coefficient was employed in this study to analyze the relative agreement between the ability estimations made in CTT and in Rasch analysis, and paired sample *t*-test was used in order to identify the absolute agreement. Finally, histogram graph was created to see the extent to which the ability estimations made through CTT and the ones made through Rasch analysis agreed in terms of distribution, and skewness and kurtosis coefficients were checked. SPSS 20.0 package programme was employed for ability estimations in CTT, for correlation analysis, paired samples *t*-test, histogram graph, and for kurtosis and skewness coefficients while FACETS package programme was used in Rasch analysis.

Correlation coefficient between ability measures in CTT and in Rasch model was found to be .95, and it was concluded that there was a high relative agreement between ability estimations made in these two theories. The findings obtained in paired samples *t*-test suggested that the ability estimations made in CTT were significantly higher than those made in Rasch analysis, and consequently, the result was considered as evidence that there were no absolute agreements between the ability scores estimated in these two theories. Finally, it was found that the ability scores estimated in Rasch analysis had more symmetrical and sharper distribution than the ones in CTT.