

# **Comparing the Scale Values Obtained from Pairwise Comparison Scaling Method and Rasch Analysis**

**Neşe Güler**

Sakarya University

**Mustafa İlhan**

Dicle University

**Gülşen Taşdelen Teker**

Hacettepe University

## **Abstract:**

*This study aims to compare the scale values obtained from pairwise comparison scaling method and Rasch Analysis. The study was conducted with 185 university students in the Spring Semester of 2015-2016 academic year. The data were collected in two stages. At the first stage, the participants were asked to make a pairwise comparison of six subjects taught in the Measurement and Evaluation course and to say which one was more difficult for them. Using the data collected in this way, scaling was performed through pairwise comparisons. At the second stage, the participants were asked to grade the six subjects as easy, medium and difficult by taking the difficulty levels of the subjects into consideration. The data obtained from this application were analyzed according to the Rasch model. It was observed that the ordering made in relation to the difficulty levels of the subjects through pairwise comparison method and through Rasch analysis was the same. Besides ordering, the closeness of the difficulty levels of the subjects were also found similar in both methods. These results demonstrate that the scaling values calculated in pairwise comparison scaling method and in Rasch analysis are compatible.*

**Keywords:** *Thurstone, scaling with pairwise comparison, Rasch analysis, interval scale*



Inönü University  
Journal of the Faculty of Education  
Vol 19, No 1, 2018  
pp. 31-48  
DOI: 10.17679/inuefd.400386

Received : 14.03.2017  
Accepted : 05.11.2017

## **Suggested Citation**

Güler, N., İlhan, M. & Taşdelen Teker, G. (2018). Comparing the Scale Values Obtained from Pairwise Comparison Scaling Method and Rasch Analysis, *Inonu University Journal of the Faculty of Education*, 19(1), 31-48. DOI: 10.17679/inuefd.400386

## EXTENDED ABSTRACT

### **Introduction**

This study aims to compare the scale values obtained from pairwise comparison scaling method and Rasch analysis. Pairwise scaling and Rasch analysis both are the methods usable in bringing the data obtained through lower order scales up to the level of interval. Therefore, this study may be said to be a study comparing two different theories which serve to the same purpose. As is commonly known, one of the most important components of scientific studies is to compare different theories and thus to determine the aspects which work and which do not work, and to exhibit the differences and similarities between them. This is also true for the branch of measurement and evaluation. Comparing theories of measurement introduced for the same purpose and thus determining their power of providing information, finding the one which is more useful and more accurate in application, and determining their similarities and differences are very important for the improvement of the science of measurement and evaluation. It is hoped that findings to be obtained in this respect would contribute to the field of measurement and evaluation as a science. Besides, considering the fact that measurement occupies a considerable place in science and that advances in many branches of science gain momentum in parallel to the development in the field of measurement, it may be said that such a study would also help in developments in different disciplines indirectly.

### **Method**

The study was conducted with 185 university students in the Spring Semester of 2015-2016 academic year. The data were collected in two stages. At the first stage, the participants were asked to make a pairwise comparison of six subjects taught in the Measurement and Evaluation course and to say which one was more difficult for them. Using the data collected in this way, scaling was performed through pairwise comparisons. At the second stage, the participants were asked to grade the six subjects as *easy*, *medium* and *difficult* by taking the difficulty levels of the subjects into consideration. The data obtained from this application were analyzed according to the Rasch model. The calculations for scaling through pairwise comparisons were made on the Microsoft Office Excel Programme, and the FACETS package programme was employed for Rasch analysis.

### **Findings**

It was observed that the ordering made in relation to the difficulty levels of the subjects through pairwise comparison method and through Rasch analysis was the same. Accordingly, it was found both in pairwise comparison scaling method and Rasch analysis that the subjects became gradually more and more difficult from the first to the last subject. Besides ordering, the closeness of the difficulty levels of the subjects were also found similar in both methods. In other words, the scale values calculated at the level of interval scale were found to be similar in terms of closeness or remoteness in both methods. For instance, subject three and subject four were very close in both methods in terms of difficulty. These results demonstrate that the scaling values calculated in pairwise comparison scaling method and in Rasch analysis are compatible. The findings obtained in the study overlap with the theoretical information and results of empirical researches about the subject.

### **Discussion & Conclusion**

Based on the results obtained it may be said that both pairwise comparison scaling technique and Rasch analysis can be used to bring the data collected through such lower level scales as ordinal scale up to the level of interval; and the scaling values to be obtained through these two methods would be compatible. Since interpretations for the scaling values about stimuli will be similar in both methods, researchers themselves can decide on the choice of methods. Yet, advantages and disadvantages of the methods should also be considered while deciding on a method. When items are presented to participants through grading instead of pairwise comparisons, data is collected in a shorter time, and participants can understand more easily how to fill in the measurement tool. Therefore, compared to pairwise comparison scaling method, it can be easier to collect the data in Rasch analysis. This difference in the process of data collection in these methods can be considered as the advantage of Rasch analysis in contrast to pairwise comparison scaling. The situation is just the opposite in the analysis of the data. Analyses regarding

pairwise comparison scaling can be performed by means of Microsoft Office Excel programme. Most of the computer programmes usable for Rasch analyses (FACETS programme for instance used in this study), on the other hand, can be reached in return for a certain amount of payment. Therefore, it may be said that pairwise comparisons scaling method would be a better choice for researchers who wish to reach scaling values at the level of interval in relation to any set of stimuli but who do not have a licenced programme through which they can do Rasch analysis. When making a choice about the method to be used, it is important to consider these differences.

# **İkili Karşılaştırmalarla Ölçekleme Yöntemi ile Rasch Analizinden Elde Edilen Ölçek Değerlerinin Karşılaştırılması**

**Neşe Güler**

Sakarya Üniversitesi

**Mustafa İlhan**

Dicle Üniversitesi

**Gülşen Taşdelen Teker**

Hacettepe Üniversitesi

## **Öz**

Bu araştırmada, ikili karşılaştırmalarla ölçekleme yöntemi ile Rasch analizinden elde edilen ölçek değerlerinin karşılaştırılması amaçlanmıştır. Araştırma 2015-2016 öğretim yılı bahar döneminde 185 üniversite öğrencisi üzerinde yürütülmüştür. Çalışmanın verileri iki adımda toplanmıştır. Birinci adımda katılımcılardan ölçme ve değerlendirme dersi kapsamında işlenen altı konuyu ikişerli karşılaştırmaları ve karşılaştırdıkları iki konudan hangisinin kendileri için daha zor olduğunu bildirmeleri istenmiştir. Bu şekilde toplanan veriler üzerinden ikili karşılaştırmalarla ölçekleme çalışması yürütülmüştür. İkinci adımda, katılımcıların aynı altı konuyu güçlük düzeyleri bakımından kolay, orta ve zor şeklinde derecelmeleri istenmiştir. Bu uygulamadan elde edilen veriler ise Rasch modeline göre analiz edilmiştir. Araştırmada, konuların güçlük düzeyleri için ikili karşılaştırmalarla ölçekleme yöntemi ve Rasch analizi sonucunda ulaşılan sıralamaların aynı olduğu görülmüştür. Sıralamanın yanı sıra konuların güçlük düzeylerine ilişkin eşit aralıklı ölçek düzeyinde hesaplanan değerlerin birbirlerine yakınlığı/uzaklığı da her iki yöntemde aynı çıkmıştır. Bu sonuçlar, ikili karşılaştırmalarla ölçekleme yönteminde ve Rasch analizinde hesaplanan ölçek değerlerinin tutarlı olduğunu ortaya koymaktadır.

**Anahtar Kelimeler:** Thurstone, ikili karşılaştırmalarla ölçekleme, Rasch analizi, eşit aralıklı ölçek.



Inönü Üniversitesi  
Eğitim Fakültesi Dergisi  
Cilt 19, Sayı 1, 2018  
ss. 31-48  
DOI: 10.17679/inuefd.400386

Gönderim Tarihi : 14.03.2017

Kabul Tarihi : 05.11.2017

## **Önerilen Atf**

Güler, N., İlhan, M. & Taşdelen Teker, G. (2018). İkili Karşılaştırmalarla Ölçekleme Yöntemi ile Rasch Analizinden Elde Edilen Ölçek Değerlerinin Karşılaştırılması. *Inönü Üniversitesi Eğitim Fakültesi Dergisi*, 19(1), 31-48. DOI: 10.17679/inuefd.400386

## GİRİŞ

Ölçme ve değerlendirme çalışmalarının tümünde geçerli ve güvenilir ölçümlere en ekonomik yoldan ulaşılması hedeflenir. Bu hedefi gerçekleştirebilmek için aynı amaçla kullanılabilen farklı yöntem ya da analizler arasındaki tutarlılığın belirlenmesi ve hangi yöntemin daha kullanışlı biçimde daha doğru sonuçlar verdiğinin saptanması gerekir. Bu sebeple, ortak bir amaç doğrultusunda işe koşulabilen farklı yöntemlerin karşılaştırılmasına yönelik araştırmalar ölçme değerlendirme alanyazınında oldukça önemli bir yer tutmaktadır. Değişen madde fonksiyonunun tespitinde kullanılan lojistik regresyon, Mantel - Haenszel, ki kare, MIMIC ve SIBTEST tekniklerinin veya kayıp veri sorununun çözümünde başvurulan seri ortalaması atama, yakın noktaların medyanının/ortalamasının alınması, liste bazında silme, doğrusal değer kestirimi ve beklenti maksimizasyonu gibi farklı yöntemlerin karşılaştırıldığı çalışmalar (Akın Arıkan, Uğurlu & Atar; Doğan & Öğretmen, 2008; Köse & Öztumur, 2014; Şahin Kürşad & Nartgün, 2015) sözü edilen araştırmalara örnek teşkil etmektedir. Bu araştırma da benzer bir düşünceyle ortaya çıkmış olup madde ölçeklemede kullanılan farklı tekniklerin karşılaştırılmasını konu edinmektedir. Bu kapsamda çalışmada ikili karşılaştırmalar yöntemi ile Rasch analizinden elde edilen ölçek değerleri arasındaki tutarlılığın incelenmesine odaklanılmış ve konu ile ilgili kuramsal bilgilere aşağıda yer verilmiştir.

### Kuramsal Çerçeve

Bir ölçme işleminden elde edilen ölçümlerin taşıdığı anlam, ölçme işlemine konu olan değişkenlere ve bu değişkenlerin ölçülmesinde kullanılan koşullara bağlı olarak farklılık gösterir. Ölçmede bu farklılığı açıklamak amacıyla ölçek kavramından yararlanır. Ölçekler, ölçümlerin verdiği bilgi düzeyi ve ölçme sonuçları üzerinde yapılabilecek matematiksel/istatistiksel işlemler bakımından sınıflama, sıralama, eşit aralıklı ve eşit oranlı olmak üzere dört başlıkta incelenir. Sınıflama ölçeğinde varlıklar ölçülen özellik açısından gösterdikleri benzerlik ve farklılıklara göre gruplara ayrılır. Bu ayırma işlemine kendi içerisinde türdeş bir yapıya sahip nesnelere aynı sınıfta yer alırken, ölçülen özellik yönüyle farklılık gösterenler farklı sınıflara yerleştirilir. Örneğin bireyleri medeni durumlarına göre bekâr ve evli şeklinde kategorilere ayırırken sınıflama ölçeğinden yararlanır. Sıralama ölçeğinde, nesnelere ya da kişiler ölçülen özelliğe sahip oluş derecesine göre sıraya dizilir. Sıralama düzeyindeki ölçek ile yapılan ölçme işlemine örnek olarak bir yüzme şampiyonasındaki yüzücülerin yarışı tamamlama sürelerine göre sıraya konulması gösterilebilir. Eşit aralıklı ölçekler, tanımlanmış (izafi) bir sıfır noktasının bulunduğu ve birimler arasındaki uzaklıkların eşit olduğu ölçek türüdür. Bir yerleşim biriminin rakımının ölçülmesi, deniz seviyesinin sıfır olarak kabul edildiği bir ölçme işlemine ifade eder ve eşit aralık düzeyindeki ölçmeye bir örnek olabilir. Eşit oranlı ölçek ise birimler arası uzaklıklarının eşit olması yönüyle eşit aralıklı ölçek ile benzerlik gösterir ancak eşit aralıklı ölçekten farklı olarak gerçek (mutlak) bir sıfır noktası bulundurmaz. Baskül kullanılarak bireylerin ağırlıklarının ölçülmesi, eşit oranlı ölçek ile gerçekleştirilen ölçme işlemine örnek olarak verilebilir.

Yukarıdaki ölçek türleri matematiksel olarak en az bilgi taşıyandan en çok bilgi taşıyana doğru sıralanmıştır. Bir başka deyişle, sınıflama ölçeğinden eşit oranlı ölçeğe doğru gidildikçe ölçme işleminin sağladığı bilgi artmakta ve ölçümler üzerinde daha fazla matematiksel işlem gerçekleştirilebilmektedir. Şöyle ki sınıflama ölçeğiyle elde edilen ölçümlere yüzde, frekans ve mod (tepe değer) dışında hiçbir matematiksel işlem uygulanamamaktadır. Sıralama ölçeğiyle toplanan veriler üzerinde yüzde, frekans ve mod hesabının yanı sıra büyüklük-küçüklük sıralamaları da yapılabilmektedir. Fakat sıralama ölçeğine ilişkin veriler toplama, çıkarma, çarpma ya da bölme gibi matematiksel işlemler yapmaya uygun değildir. Eşit aralıklı ölçek düzeyindeki veriler sıralama ölçeğindeki verilere uygulanabilen işlemlere ek olarak toplama ve çıkarma işlemlerini de gerçekleştirmeyi mümkün kılmaktadır. Ancak gerçek bir sıfır noktasının bulunmamasından dolayı eşit aralıklı ölçek düzeyindeki verilere çarpma ve bölme işlemi uygulanamamaktadır. Eşit oranlı ölçekte ise ölçümler üzerinde bütün matematiksel işlemler yapılabilmektedir. Sınıflama ve sıralama ölçeğiyle toplanan verilere uygulanabilecek matematiksel ve istatistiksel işlemler kısıtlı olduğundan ölçme çalışmalarında eşit aralıklı ya da eşit oranlı ölçümlere ulaşılması hedeflenir (Turgut & Baykul, 1992). Bununla birlikte, sosyal bilimlerde elde edilen ölçümler genellikle sınıflama ve sıralama düzeyinde yer alır. Ölçme sonuçları sınıflama ve sıralama düzeyinde olmasına rağmen en azından eşit aralık düzeyindeki ölçümlere ulaşılacak istenmesi, araştırmacıları daha az bilgi veren ölçme sonuçlarını eşit aralıklı veri haline getirebilecek yöntem arayışlarına sevk etmiştir (Baykul, 1996). Bu arayışlar sonucunda ortaya çıkan yöntemlerden biri ölçekleme teknikleridir.

### **Ölçekleme Teknikleri**

Ölçekleme, değişkenler arasındaki farklılıkları daha net bir biçimde ortaya koymak ve ölçümler üzerinde yapılabilecek matematiksel işlemleri zenginleştirmek için sınıflama ve sıralama düzeyindeki verilerin eşit aralıklı ölçek düzeyine getirilmesini ifade eder. Ölçekleme sürecinde kullanılan teknikler denek tiplerine dayalı yaklaşımlar ve yargıcı kararlarına dayalı yaklaşımlar olmak üzere iki başlıkta incelenir. Denek tepkilerine dayalı yaklaşımlar cevaplayıcı merkezlidir (subject centered) ve cevapların ölçeklenmesi esasına dayanır (Torgerson, 1958). Bu yaklaşımda maddelere verdiği tepkilere dayalı olarak her bir birey ölçek üzerinde farklı bir yere yerleştirilir (Crocker & Algina, 1986). Denek tiplerine dayalı yaklaşımlar, dereceleme toplamlarıyla ölçekleme ve çok boyutlu ölçekleme tekniklerinden oluşur (Turgut & Baykul, 1992).

Yargıcı kararlarına dayalı yaklaşımlar ise denek tiplerine dayalı yaklaşımların aksine cevaplayıcı değil; uyarıcı merkezlidir (stimulus centered). Uyarıcı merkezli yaklaşımların temelinde her bir uyarıcının psikolojik ölçek üzerindeki yerinin gözlemci yargılarına dayalı olarak belirlenmesi yer alır (Kan, 2008a; Arık & Kutlu, 2013). Bir diğer ifadeyle, denek tepkilerine dayalı yaklaşımda cevapların ölçeklenmesine odaklanılırken; yargıcı kararlarına dayalı yaklaşımda madde ya da uyarıcıların ölçeklenmesi esastır (Torgerson, 1958). Yargıcı kararlarına dayalı yaklaşımlar; ikili karşılaştırmalarla ölçekleme, sınıflama yargılarıyla ölçekleme, sıralama yargılarıyla ölçekleme ve mutlak yargılarla ölçekleme olmak üzere dört başlıkta ele alınmaktadır (Kan, 2008b). Bu araştırmada, sıralanan bu dört yaklaşımdan ikili karşılaştırmalarla ölçekleme yöntemi üzerinde durulmuştur.

### **İkili Karşılaştırmalarla Ölçekleme**

İkili karşılaştırmalarla ölçekleme yöntemi Thurstone (1927) tarafından geliştirilmiştir. Bu yöntem başlangıçta tutum cümlelerinin ölçeklenmesinde kullanılmıştır. Bununla birlikte, yöntem tutumların ölçülmesiyle sınırlı olmayıp geniş bir uygulama alanına sahiptir (Turgut & Baykul, 1992). Herhangi bir iş grubundaki bireylerin mesleki sorunlarının tespiti (Anıl, Ertuna & Uysal, 2015), bir öğretim programına başvuran öğrencilerde aranan özelliklerin belirlenmesi (Güler & Anıl, 2009), bireylerin sosyal aktivite tercihleri (Polat & Göksel, 2014) ya da öğrencilerin üniversite/okul tercihleri (Ertoprak, 2015) gibi cevaplayıcıların uyarıcıları ikiye bölerek değerlendirilebileceği her durumda ikili karşılaştırmalarla ölçekleme yöntemi kullanılabilir (Anıl & Güler, 2006). Bu ölçekleme yönteminde katılımcılardan, kendilerine sunulan uyarıcı çiftlerinden birini diğerine tercih etmeleri istenir. Daha açık bir ifadeyle, gözlemcilerden karşılaştırdıkları iki uyarıcıdan hangisinin "daha öncelikli", "daha iyi", "daha olumsuz", "daha zor" vb. olduğuyla ilgili mutlak bir seçim yapmaları beklenir. Ayırt etmeme yargılarına ise izin verilmez (Turgut & Baykul, 1992). Sunulan uyarıcılar gözlemciler tarafından ikiye bölerek karşılaştırıldıktan sonra her bir uyarıcıya ilişkin ölçek değerlerinin hesaplanmasına yönelik istatistiksel işlemlere başlanır. Bu kapsamda takip edilen adımlar; i) frekans matrisinin oluşturulması, ii) oranlar matrisinin hesaplanması, iii) birim normal sapmalar matrisinin elde edilmesi, iv) birim normal sapmalar matrisindeki her bir sütunun ortalamasının bulunması ve v) bulunan ortalamalara ölçek değerlerinin başlangıç noktasını sıfıra kaydıracak bir öteleme uygulanması şeklinde sıralanmaktadır (Erkuş, 2012; Tekindal, 2009). Bu adımların tamamlanmasıyla ikili karşılaştırmalarla toplanan veriler eşit aralıklı ölçek düzeyine getirilmiş olur. Ölçekleme tekniklerinin yanı sıra sıralama ölçeğindeki verileri eşit aralık düzeyine getirmek amacıyla başvurulabilecek bir diğer yöntem de Rasch modelidir.

### **Rasch Modeli**

Rasch modeli, bir parametrelili Madde Tepki Kuramı (MTK) modeli olarak da bilinmektedir (Baker, 2001). Yalnızca madde güçlük parametresi içeren bu model ilk olarak doğru yanlış şeklinde puanlanabilen iki kategorili (dichotomous) maddeler için geliştirilmiştir (Rasch, 1960). Ancak ilerleyen dönemlerde modelin uygulamaları iki kategorili maddeler ile sınırlı kalmamıştır. Andrich'in (1978a) sıralama ölçekli modeli (rating scale model) geliştirilmesiyle Rasch modeli Likert tipi ölçek verilerinin analizinde de kullanılabilir hale gelmiştir. Andrich'in (1978a) çalışmalarından kısa bir süre sonra Masters (1982) kısmi puanlama modelini (partial credit model) ileri sürmüştür. Bu gelişmeyle birlikte Rasch modeli açık uçlu soruların analizinde de kullanılmaya başlanmıştır. Rasch modelinin gelişimine katkıda bulunan bir diğer isim Linacre'dir. Linacre'nin (1989) kurama katkısı, Rasch modelini bireyler ve maddelerin yanı sıra farklı değişkenlik kaynaklarını içerecek şekilde genişletmesi olmuştur. Gerek iki kategorili maddeler için geliştirilen temel Rasch modelinde, gerek Andrich tarafından geliştirilen sıralama ölçekli modelde gerekse de Masters tarafından ileri sürülen kısmi puanlama modelinde ölçme sonuçlarını etkileyebilecek değişkenlik kaynakları, bireyler ve maddeler ile sınırlıdır. Linacre'nin (1989) çok yüzeyli Rasch modelini (many facet Rasch model) geliştirmesiyle bireyler ve maddelerin yanı sıra puanlayıcılar gibi ölçme sonuçları üzerinde etkili olabilen başka değişkenlik kaynakları da modele dâhil edilmiştir.

Rasch analizinin ölçme değerlendirme süreçleri açısından sağladığı çeşitli avantajlar bulunmaktadır. Öncelikle, MTK kapsamında yer alan bir model olması dolayısıyla MTK'ya ilişkin bireylerin yetenek düzeylerinin madde parametrelerinden bağımsız hesaplanabilmesi ve madde parametrelerinin bireylerin yetenek düzeylerinden bağımsız kestirilebilmesi varsayımları Rasch modeli için de geçerlidir (DeMars, 2010). Rasch modelinin bir diğer avantajı; ölçme işlemine dâhil olan tüm değişkenlik kaynaklarını ortak bir ölçek üzerine yerleştirerek bu değişkenlik kaynakları arasında karşılaştırma yapmayı kolaylaştırmasıdır (O'Neill & Lunz, 1996; Schaefer, 2008). Rasch modelinin en önemli avantajlarından biri de sıralama ölçeğindeki orijinal verileri eşit aralık düzeyine getirmesidir. Örneğin, Likert tipi ölçekler ile toplanan orijinal veriler sıralama düzeyinde olmasına rağmen, bu veriler Rasch modeline göre analiz edildiğinde eşit aralıklı logit ölçeğine dönüştürülmektedir (Wright, & Masters, 1982).

### **Araştırmanın Amacı ve Önemi**

Bu çalışmada ikili karşılaştırmalarla ölçekleme çalışması ile Rasch analizinden elde edilen ölçek değerlerinin karşılaştırılması amaçlanmıştır. Hem ikili karşılaştırmalarla ölçekleme yöntemi hem de Rasch analizi; sıralama ölçeği ile elde edilen verileri eşit aralık düzeyine getirmek amacıyla başvurulabilen yöntemlerdir. Dolayısıyla bu araştırmanın aynı amaca hizmet eden farklı ölçme kuramlarının karşılaştırılmasına yönelik bir çalışma olduğu söylenebilir. Benzer amaçla ileri sürülen farklı kuramların karşılaştırılıp birbirlerine göre üstünlük ve zayıflıklarının belirlenmesi, uygulamada daha kullanışlı ve doğru olanın saptanması, tutarlılık gösteren ya da örtüşmeyen yanlarının ortaya konulması bilimsel çalışmaların önemli parçalarından biridir (Atılğan, 2004; Doğan, 2002). Bu anlamda çalışmada ulaşılabilecek bulguların bir bilim dalı olarak ölçme alanının gelişmesine katkı sağlayacağı umulmaktadır. Ayrıca bilimde ölçmenin önemli bir yere sahip olduğu ve birçok bilim dalındaki ilerlemelerin ölçme alanındaki gelişmelere paralel olarak hız kazandığı (Baykul, 2010) dikkate alındığında ölçme değerlendirme bilim dalının gelişmesine katkı sağlayacak böyle bir çalışmanın dolaylı olarak farklı disiplinlerdeki gelişmelere de aracılık etmesi beklenmektedir.

Alanyazına bakıldığında ölçeklemede kullanılan farklı yaklaşımların karşılaştırılmasına yönelik çalışmaların (Andrich, 1978; Engelhard, 1984; Jansen, 1984; Wainer, Fairbank & Hough, 1978) bulunduğu görülmektedir. Andrich'in (1978) yaptığı çalışmada kuramsal bir bakış açısıyla madde ölçeklemede Thurstone ve Rasch yaklaşımları arasındaki ilişki üzerinde durulmuştur. Wainer, Fairbank ve Hough (1978) tarafından yapılan çalışmada ikili karşılaştırmalarla ölçekleme yöntemi ile Rasch analizi ampirik olarak karşılaştırılmıştır. Jansen'in (1984) yaptığı çalışma, farklı ölçekleme tekniklerinin kuramsal olarak karşılaştırıldığı bir başka çalışmadır. Jansen (1984) kuramsal bir çerçevede Thurstone, Coombs ve Rasch ölçekleme yöntemleri arasındaki ilişkiyi ele almıştır. Engelhard (1984) tarafından yürütülen çalışmada ise psikoloji ve eğitim alanındaki testlerin ölçeklenmesinde kullanılan Thorndike, Thurstone ve Rasch ölçekleme yöntemleri hem kuramsal hem de uygulamalı olarak karşılaştırılmıştır.

Bu araştırma çeşitli açılardan yukarıda sıralanan çalışmalardan farklılık göstermektedir. Öncelikle bu çalışmada, ikili karşılaştırmalarla ölçekleme tekniği ile Rasch analizinden elde edilen ölçek değerleri arasındaki tutarlılık ampirik olarak test edilmiştir. Araştırma bu yönüyle Andrich (1978) ve Jansen'in (1984) sözü edilen iki yöntemi kuramsal temelleriyle karşılaştırdığı çalışmalardan ayrılmaktadır. Araştırmayı Wainer vd. (1978) ile Engelhard'ın (1984) çalışmasından farklı kılan özellikler ise şu şekilde açıklanabilir. Wainer vd. (1978) tarafından yapılan çalışmada, 95 maddelik bir veri toplama aracı dört farklı gruptan toplam 355 katılımcıya uygulanmıştır. Veri toplama aracındaki 95 maddeden, model ile yeterli uyum gösteren (hem örneklemin genelinde hem de dört grubun her birinde) ve dört farklı grup boyunca ölçek değerleri arasında anlamlı bir farklılık bulunmayan 51 tanesi için Rasch analizine ait ölçek değerleri rapor edilmiştir. Sonrasında, Rasch analizinde dikkate alınan 51 maddeden beş tanesi 355 katılımcının 74'üne ikili karşılaştırmalar şeklinde sunulmuştur. Bu veriler üzerinden ikili karşılaştırmalarla ölçekleme işlemi gerçekleştirilip hesaplanan ölçek değerleri Rasch analizinden elde edilen ölçek değerleriyle karşılaştırılmıştır. Buna göre, Wainer vd.'nin (1978) çalışmasında Rasch analizi ile ikili karşılaştırmalarla ölçekleme çalışmasına dâhil olan katılımcılar ve iki yöntemle ilişkin verilerin elde edilmesine kullanılan uyarıcılar tam olarak eş değer değildir. Wainer vd.'nin (1978) çalışmasının aksine, bu çalışmada ikili karşılaştırmalarla ölçekleme tekniği ile Rasch analizine ilişkin veriler, aynı çalışma grubuna aynı maddelerin uygulanması ile elde edilmiştir. Bu özelliği, araştırmayı Wainer vd.'nin (1978) çalışmasından farklı kılmakta ve böylesi bir farklılığın iki ayrı analiz tekniğinin karşılaştırılmasının amaçlandığı bir çalışma için önemli olduğu düşünülmektedir. Ayrıca, ölçeklemede kullanılan değişik yöntemler arasındaki benzerlik veya farklılıklarda veri setinden kaynaklı etkiler de mümkün olduğundan (Engelhard, 1984), ikili karşılaştırmalarla ölçekleme yöntemi ve Rasch analizinin karşılaştırılmasının amaçlandığı bu araştırmanın

Wainer vd.'nin (1978) çalışmasında ulaşılan bulguların genellenebilirliğine katkı sunması beklenmektedir. Bu araştırmayı Engelhard'ın (1984) çalışmasından ayıran özellik ise Engelhard (1984) tarafından yapılan çalışmada Thurstone'un mutlak yargılarla ölçekleme tekniği kullanılırken, bu çalışmada Thurstone'un ikili karşılaştırmalarla ölçekleme tekniğinin kullanılacak olmasıdır. Sıralanan tüm bu hususlar göz önüne alındığında, araştırmamızın alanyazına katkı sağlayacağı söylenebilir.

## YÖNTEM

### Araştırma Modeli

İkili karşılaştırmalarla ölçekleme çalışmasına ait sonuçlar ile Rasch analizi çıktılarının karşılaştırılmasının amaçlandığı bu çalışma temel bir araştırma niteliğindedir. Temel araştırmaların odak noktasında yeni bir kuramın ileri sürülmesi veya mevcut kuramların geliştirilmesi yer alır (Powell, 1997). Gerçek yaşam problemlerine cevap bulma düşüncesiyle ortaya çıkan uygulamalı araştırmaların aksine temel araştırmaların uygulamaya yönelme endişesi bulunmamaktadır. Ancak bu durum temel araştırmaların uygulamaya yönelik herhangi bir katkı sağlamayacağı anlamına gelmemektedir (Hall, 2008). Uygulamalı araştırmaların büyük bir kısmı temel araştırmalarda ulaşılan bulgular üzerine inşa edildiğinden temel araştırmaların uygulamaya dönük çalışmalara katkıları da olmaktadır (Jackson, 2008; Johnson & Christensen, 2014).

### Çalışma Grubu

Araştırma, Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi ile Sakarya Üniversitesi Hendek Eğitim Fakültesi'nde ölçme değerlendirme dersine devam eden 185 üniversite öğrencisi üzerinde yürütülmüştür. Bu öğrencilerin 89'u (% 48.11) Dicle Üniversitesi'ne kayıtlı iken, geriye kalan 96 (% 51.89) öğrenci Sakarya Üniversitesi'ne kayıtlıdır. Çalışma grubunda; Din Kültürü ve Ahlak Bilgisi [41 kişi (%22.16)], İlköğretim Matematik [48 kişi (%25.95)], Okul Öncesi [42 kişi (%22.70)] ve Sosyal Bilgiler Öğretmenliği [54 kişi (%29.19)] olmak üzere dört farklı programdan öğrenciler yer almıştır. Turgut ve Baykul (1992) ikili karşılaştırmalarla ölçekleme çalışması için 100 gözlemcinin yer aldığı bir örneklemin yeterli olduğunu belirtmektedir. Benzer şekilde, Rasch analizinde 100 ile 200 kişiden toplanan veriler parametre kestirimleri için yeterli görülmektedir (Demars, 2010). Buna göre, çalışma grubundaki katılımcı sayısının bu tür bir araştırma için uygun olduğu söylenebilir.

### Veri Toplama Araçları ve Verilerin Toplanması

Araştırmada, katılımcıların herhangi bir konu hakkındaki düşüncelerinin belirlenmesi ile ilgilenilmediğinden kullanılacak veri toplama araçlarının içeriğinin ne olduğu çalışma sonuçlarını etkileyebilecek bir faktör değildir. Dikkat edilmesi gereken temel nokta, ikili karşılaştırmalarla ölçekleme yöntemi ile Rasch analizine ilişkin verilerin toplanmasında kullanılacak ölçme araçlarında aynı ifadelerden oluşan maddelerin yer almasıdır. Bu nedenle, veri toplama araçlarının içeriğine karar verirken yalnızca ölçme araçlarında yer verilen içeriğe uygun bir örneklemin araştırmacılar için kolay ulaşılabilir olması esas alınmıştır. Bu doğrultuda veri toplama araçlarında ölçme değerlendirme dersi konularının güçlük düzeyleriyle ilgili bir içerik benimsenmiştir.

Veri toplama araçlarını hazırlamak için ilk olarak lisans düzeyinde okutulan ölçme değerlendirme ders kitapları (Atılğan, 2013; Bahar, Nartgün, Durmuş & Bıçak, 2010; Başol, 2013; Çıkrıkçı Demirtaşlı, 2012; Gömlüksiz & Erkan, 2010; Güler, 2012; Tan, 2015) incelenmiştir. Bu incelemeyle derse ilişkin konuların hangi başlıklar altında toplanabileceği belirlenmeye çalışılmıştır. Kitaplardaki konu başlıkları gözden geçirildiğinde ölçme değerlendirme dersi konularına ilişkin farklı kitaplarda değişik sınıflandırmalar yapıldığı saptanmıştır. Bu sebeple, ölçme değerlendirme dersi konularının hangi başlıklar altında ifade edilmesi gerektiğine karar verirken, incelenen kitaplarda yer alan ortak başlıklar ve en sık kullanılan sınıflandırmalar esas alınmıştır. Örneğin, incelenen yedi kitabın beşinde ölçme değerlendirme dersinin ilk konusu ölçme değerlendirmede temel kavramlar olarak geçmektedir. Dolayısıyla, dersin birinci konusu için ölçme değerlendirmeyle ilgili temel kavramlar başlığı uygun görülmüştür. Kitapların büyük bir kısmında temel kavramların ardından ölçme ve değerlendirme açısından hata kavramı ile hatanın türleri (sabit, sistematik ve tesadüfi) ele alındığından, dersin ikinci konusu için ölçmede hata ifadesi kullanılmıştır. Bu şekilde bir yol takip edilerek ölçme değerlendirme dersi konuları Tablo 1'de sunulan altı başlık altında toplanmıştır.



Tablo 1.

*İşleniş sırasına göre ölçme değerlendirme dersinin temel konuları*

Konu Numaraları	Konu Başlıkları
Konu 1	Ölçme değerlendirmeye ilgili temel kavramlar
Konu 2	Ölçmede hata
Konu 3	Ölçme araçlarında bulunması gereken nitelikler
Konu 4	Eğitimde kullanılan ölçme araç ve yöntemleri
Konu 5	Test geliştirme süreci ve madde analizi
Konu 6	Test puanları üzerinde istatistiksel işlemler

Araştırma kapsamında, Tablo 1’de belirtilen konular iki farklı form şeklinde katılımcılara sunulmuştur. Bu formlardan ilki, ölçme değerlendirme dersi konularının ikili karşılaştırmalar halinde düzenlenmesiyle oluşturulmuştur. Dolayısıyla araştırmada kullanılan veri toplama araçlarının ilkinde, Tablo 1’deki altı konunun ikili kombinasyonlarından oluşan 15 madde yer almıştır. Katılımcılardan karşılaştırılan iki konudan hangisinde daha fazla zorlandıklarını bildirmeleri istenmiştir. Veri toplama aracının başına bir uyarı eklenerek sunulan ikililerden mutlaka birinin tercih edilmesi gerektiği belirtilmiştir. Buna rağmen, yanıtız bırakılan ikili karşılaştırma içeren ölçme araçları kapsam dışı tutulmuştur.

Araştırmada kullanılan ikinci veri toplama aracında ise ölçme değerlendirme dersi konuları *kolay* (1), *orta* (2) ve *zor* (3) şeklinde üçlü bir dereceleme ile katılımcılara uygulanmıştır. Katılımcılardan sunulan konuların güçlük düzeylerini dikkate alarak, uygun olduğunu düşündüğü seçeneği işaretlemeleri talep edilmiştir. İkili karşılaştırmalara ait ölçek formunda olduğu gibi, Likert tipi üçlü derecelemenin kullanıldığı bu ölçek formunda da yanıtız bırakılan maddelerin yer aldığı ölçme araçları veri dosyasına dâhil edilmemiştir. Ölçme araçlarının her ikisi de sınıf ortamında öğrencilere uygulanmıştır. Uygulamalar 2015-2016 öğretim yılı bahar dönemi sonunda bizzat araştırmacılar tarafından gerçekleştirilmiştir. Uygulama sırasında araştırmaya katılımın zorunlu olmadığı öğrencilere hatırlatılarak çalışma grubunun sadece gönüllü öğrencilerden oluşması sağlanmıştır. İkili karşılaştırmalar ölçek formu, ölçme ve değerlendirme dersinin son konusu olan standart puanların işlendiği dersin başında, Likert tipi üçlü derecelemenin kullanıldığı ölçek formu ise aynı dersin sonunda öğrencilere uygulanmıştır. Bu sebeptendir ki; standart puanlar konusu çalışmaya dâhil edilmemiştir.

**Veri Analizi**

Araştırmada ikili karşılaştırmalar ölçek formu ile toplanan verilerin analizi için Thurstone’un karşılaştırmalı yargı kanununun beşinci hal denklemine başvurulmuştur. Bu doğrultuda öncelikle çalışma grubundaki öğrencilerin ölçme değerlendirme dersi konularının güçlük düzeyine ilişkin yaptıkları karşılaştırmalara ait frekans değerleri hesaplanmıştır. Hesaplanan bu değerler kullanılarak frekans matrisi oluşturulmuştur. Ardından frekans matrisindeki her bir değer çalışmadaki toplam katılımcı sayısına bölünerek oranlar matrisine ulaşılmıştır. Sonrasında, oranlar matrisinin elemanlarına karşılık gelen standart Z puanları belirlenerek birim normal sapmalar matrisi elde edilmiştir. Bu işlemi takiben birim normal sapmalar matrisinin en alt satırına, her bir sütuna ait değerlerin ortalamasına karşılık gelen bir satır eklenmiş ve bu sayede ölçek değerleri hesaplanmıştır. Ölçekleme çalışmasının son basamağında en küçük ölçek değeri sıfır olacak şekilde bir öteleme yapılmış ve ölçek değeri sıfıra daha yakın olan konular, katılımcıların daha fazla zorlandıkları konular olarak yorumlanmıştır. Araştırmada, ikili karşılaştırmalarla ölçekleme çalışmasına ilişkin bütün hesaplamalar Microsoft Office Excel programı ile gerçekleştirilmiştir.

Çalışmada, kolay-orta-zor şeklinde Likert tipi üçlü derecelemenin kullanıldığı ölçek formu ile elde edilen veriler ise Rasch modeline göre analiz edilmiştir. FACETS paket programından (Linacre, 2012) yararlanılarak gerçekleştirilen analizde *öğrenciler ve ölçme konuları* şeklinde iki yüzeyli bir desen söz konusu olmuştur. İkili karşılaştırmalarla ölçekleme işleminin gerçekleştirilmesi ve Rasch analizinin uygulanmasının ardından ölçme değerlendirme dersi konularının güçlük düzeyleri için bu iki analizde rapor edilen ölçek değerlerinin ne derece tutarlı olduğu kontrol edilmiştir.

## BULGULAR

Bu bölümde araştırmadan elde edilen bulgulara yer verilmiştir. Bulgular, iki başlık altında sunulmuştur. Birinci başlıkta ikili karşılaştırmalarla ölçekleme çalışmasından elde edilen bulgular özetlenmiş; ikinci başlıkta ise Rasch analizi sonucunda ulaşılan bulgulara değinilmiştir.

### İkili Karşılaştırmalarla Ölçekleme Çalışmasına Ait Bulgular

Araştırmada öncelikle her bir öğrenciden altı konuyu (uyarıcıyı) ikiyeşerli olarak karşılaştırmaları ve karşılaştırılan iki konudan daha zor olanını belirtmeleri istenmiştir. Elde edilen veriler üzerinden her bir konuya ait frekans değerleri saptanarak Tablo 2’de yer alan Frekans Matrisi elde edilmiştir. Tablo 2’nin her bir hücresi  $S_j > S_i$  olmak üzere  $f_{ij}$  frekans değerlerini göstermektedir. Bu matriste esas köşegenine göre birbirinin simetriği olan değerlerin toplamı çalışma grubundaki öğrenci sayısına eşittir. Ayrıca matrisin esas köşegeni üzerinde yer alan hücrelerde herhangi bir frekans değerinin bulunmadığı gözlenmektedir. Bunun nedeni, esas köşegen üzerindeki hücrelerin bir uyarıcının kendisi ile karşılaştırılması anlamını taşımasıdır. Böyle bir karşılaştırma yapılamayacağından esas köşegen üzerindeki hücrelerde herhangi bir değer yer almamaktadır.

Tablo 2.

*Frekans Matrisi\**

Uyarıcılar	U1	U2	U3	U4	U5	U6
U1 (Temel Kavramlar)	–	71	21	25	13	6
U2 (Ölçmede Hata)	114	–	78	58	11	13
U3 (Ölçme Ar. Bul. Ger. Nitelikler)	164	107	–	81	39	19
U4 (Ölçme Araç ve Yöntemleri)	160	127	104	–	37	30
U5 (Test Geliştirme ve Madde Analizi)	172	174	146	148	–	46
U6 (Test Puanı İstatistikleri)	179	172	166	155	139	–

\* Tabloda kullanılan *U* harfi uyarıcı ifadesini temsi etmektedir.

Frekans (F) matrisi oluşturulduktan sonra bu matristeki hücre değerleri çalışmadaki toplam kişi sayısına ( $N=185$ ) bölünerek, Tablo 3’teki oranlar matrisi elde edilmiştir. Bir başka deyişle, oranlar matrisinin (P) her bir elemanı, frekans matrisindeki hücre değerlerinin 185’e bölümüne karşılık gelmektedir. Oranlar matrisinde esas köşegene göre simetrik olan elemanların toplamı daima 1’e eşittir.

Tablo 3.

*Oranlar Matrisi*

Uyarıcılar	U1	U2	U3	U4	U5	U6
U1 (Temel Kavramlar)	–	0.38	0.11	0.14	0.07	0.03
U2 (Ölçmede Hata)	0.62	–	0.42	0.31	0.06	0.07
U3 (Ölçme Ar. Bul. Ger. Nitelikler)	0.89	0.58	–	0.44	0.21	0.10
U4 (Ölçme Araç ve Yöntemleri)	0.86	0.69	0.56	–	0.20	0.16
U5 (Test Geliştirme ve Madde Analizi)	0.93	0.94	0.79	0.80	–	0.25
U6 (Test Puanı İstatistikleri)	0.97	0.93	0.90	0.84	0.75	–

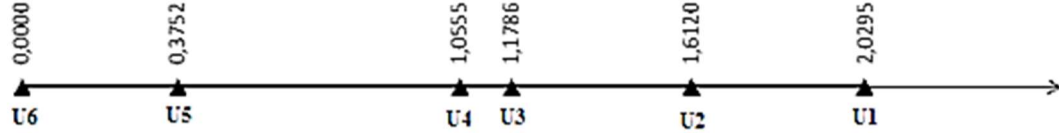
Oranlar matrisinin ardından Excel’de Normsters işlevi kullanılarak Tablo 3’teki hücre değerlerine karşılık gelen standart değerler (Z) belirlenmiştir. Böylece Tablo 4’te verilen birim normal sapmalar matrisine (Z matrisi) ulaşılmıştır. Tablo 4’ten anlaşılacağı üzere, Z matrisinde esas köşegene göre simetrik olan hücre elemanları işaretçe zıt yönlüdür ve mutlak değerce birbirine eşittir.

Tablo 4.

Birim normal sapmalar matrisi

Uyarıcılar	U1	U2	U3	U4	U5	U6
U1 (Temel Kavramlar)	–	-0.30	-1.21	-1.10	-1.47	-1.85
U2 (Ölçmede Hata)	0.30	–	-0.20	-0.49	-1.56	-1.47
U3 (Ölçme Ar. Bul. Ger. Nitelikler)	1.21	0.20	–	-0.16	-0.80	-1.27
U4 (Ölçme Araç ve Yöntemleri)	1.10	0.49	0.16	–	-0.84	-0.99
U5 (Test Geliştirme ve Madde Analizi)	1.47	1.56	0.80	0.84	–	-0.68
U6 (Test Puanı İstatistikleri)	1.85	1.47	1.27	0.99	0.68	–
$\Sigma Z_j$	5.93	3.42	0.82	0.08	-3.99	-6.25
$Z_{jort}$	0.99	0.57	0.14	0.01	-0.66	-1.04
Ölçek ( $s_j$ )	2.0295	1.6120	1.1786	1.0555	0.3752	0.0000

Tablo 4'teki  $\Sigma Z_j$  satırı, her bir sütundaki elemanların toplamına eşittir.  $\Sigma Z_j$  satırındaki değerlerin karşılaştırılan uyarıcı sayısı olan altıya (6) bölünmesiyle  $Z_{jort}$  satırındaki değerler hesaplanmıştır. Tablo 4'e göre  $Z_{jort}$  satırındaki en küçük değer -1.04'tür. Eksenin başlangıç noktasını sıfıra kaydırmak için tüm  $Z_{jort}$  değerlerine en küçük z'nin mutlak değeri olan 1.0418 eklenmiştir. Bu sayede ölçek değerleri elde edilmiştir. Hesaplanan ölçek değerleri, Şekil 1'deki sayı doğrusu üzerinde sunulmuştur.



Şekil 1. Uyarıcılara ilişkin ölçek değerlerinin sayı doğrusu üzerindeki gösterimi

Şekil 1'deki sayı doğrusu üzerinde yer alan uyarıcı değerlerine göre, ölçme ve değerlendirme dersi kapsamında işlenen konular güçlük düzeyleri açısından Tablo 5'te gösterildiği gibi sıralanmaktadır.

Tablo 5.

Ölçme ve değerlendirme konularının ölçek değerleri ve sıraları

Konular	Ölçek Değerleri	Uyarıcı Sıraları
6. Test Puanları İstatistikleri	0.0000	1
5. Test Geliştirme ve Madde Analizi	0.3752	2
4. Ölçme Araç ve Yöntemleri	1.0555	3
3. Ölçme Araç, Bul. Gerek. Nitelikler	1.1786	4
2. Ölçmede Hata	1.6120	5
1. Temel Kavramlar	2.0295	6

Tablo 5'teki en küçük ölçek değeri öğrencilerin en çok zorlandıkları konuyu gösterirken; en yüksek ölçek değeri öğrenciler tarafından en kolay olduğu düşünülen konuyu temsil etmektedir. Buna göre, *ölçme ve değerlendirmede temel kavramlar* öğrencilerin en kolay olduğunu düşündükleri konudur. *Test puanları üzerinde istatistiksel işlemler* ise öğrencilerin en çok zorlandıkları konu olarak belirlenmiştir. Üçüncü (*Ölçme araçlarında bulunması gereken nitelikler*) ve dördüncü konu (*Ölçme araç ve yöntemleri*) güçlük düzeyleri açısından birbirine en yakın konular olarak tespit edilmiştir. Tablo 5'te görüldüğü gibi ölçme ve değerlendirme dersi kapsamında yer alan konuların işleniş sırası ile konuların kolaydan zora doğru sıralanışı aynıdır.

### Rasch Analizine İlişkin Bulgular

Çalışmanın ikinci aşamasında, öğrencilerden güçlük düzeylerini dikkate alarak ölçme ve değerlendirme dersinin altı temel konusunu *kolay* (1), *orta* (2) ve *zor* (3) şeklinde değerlendirmeleri istenmiştir. Bu şekilde toplanan veriler üzerinden Rasch analizi uygulanmıştır. Rasch analizi sonucunda rapor edilen logit cetvel Şekil 2'de sunulmuştur.

Measr	+Birey	+KONULAR	Ölçek
5	+	+	(3)
4	+	+	
3	+	6	
2	+	5	
1	+	+	
0	*	*	2 *
-1	+	3 4	
-2	+	2	
-3	+	1	
-4	+	+	
-5	+	+	(1)
Measr	* = 4	+KONULAR	Ölçek

Şekil 2. Rasch analizine ilişkin logit cetvel

Şekil 2'deki logit cetvelde; öğrenciler, konular ve ölçme işleminde kullanılan üçlü derecelemeye sahip ölçek bir arada görülebilmektedir. Logit cetvelde, öğrencilerin yer aldığı sütun incelendiğinde -5'ten +5'e doğru konuları en kolay olarak belirten öğrencilerden en zor olarak belirten öğrencilere doğru bir sıralama vardır. Öğrenci sayısı fazla olduğu için Şekil 2'de her bir öğrenci tek tek gösterilememiştir. Bunun yerine her yıldızın dört, her noktanın ise bir öğrenciyi temsil ettiği bir gösterim kullanılmıştır. Konuların yer aldığı sütunda ise -5'e en yakın konu, en kolay konuyu ifade ederken; +5'e yaklaştıkça konuların güçlük düzeyi artmaktadır. Buna göre, altıncı konu öğrencilerin en zor olarak gördüğü konudur. Öğrencilerin en kolay olduğunu düşündüğü konu ise birinci konudur. Üçüncü ve dördüncü konular öğrencilerin güçlük açısından birbirine en yakın buldukları konulardır. Logit cetvelin en son sütununda ölçme işleminde kullanılan üçlü derecelemeye ait sonuçlar bulunmaktadır. Derecelemeyle ilgili kolay (1), orta (2) ve zor (3) kategorilerinin logit cetvel boyunca geniş bir aralığa sahip olması, üçlü derecelemenin etkin çalıştığına işaret etmektedir. Logit cetvelin ardından öğrencilere ilişkin ölçüm raporları incelenmiş ve sonuçlar Tablo 6'da sunulmuştur.

Tablo 6.

*Öğrencilere ilişkin ölçüm raporu*

	Ölçüm	Standart Hata	Uyum-içi	Uyum-dışı
Ortalama	0.02	0.89	0.98	1.02
Standart Sapma	1.63	0.12	0.67	1.12
Güvenirlilik = 0.70	Ayrırma oranı = 1.53	Ki-kare = 453.4	sd = 184	p = 0.00

Tablo 6'ya göre uyum-içi (INFIT) ve uyum-dışı (OUTFIT) istatistiklerine ait ortalamalar sırasıyla 0.98 ve 1.02'dir. Uyum istatistikleri model ile veri arasındaki uyum hakkında bilgi vermektedir. Uyum istatistiklerine ilişkin istenen değer 1.00'dir. Bu değer, model ile veri arasında mükemmel bir uyum olduğunu yansıtır. Ancak gerçek ölçme koşullarında mükemmel bir uyuma rastlanma olasılığı oldukça düşüktür (Brentari & Golia, 2008). Bundan dolayı model ile veri arasındaki uyum hakkında bir değerlendirme yaparken mükemmel bir uyum aramak yerine kabul edilebilir uyum olup olmadığının test edilmesi önerilmektedir.

Model ile veri arasındaki kabul edilebilir uyuma ilişkin farklı araştırmacılar değişik ölçütleri sürmüştür. Lunz, Wright ve Linacre'ye (1990) göre uyum istatistiklerine ait kabul edilebilir değerler 0.6 ile 1.5 arasında değişmektedir. Lunz ve Stahl (1990) ile Wright ve Linacre (1994), 0.5 ile 1.5 arasında kalan değerlerin uyum istatistikleri için kabul edilebilir olduğunu ifade etmektedir. Myford ve Wolfe (2003) ise 2'ye kadar olan uyum istatistiklerini kabul edilebilir olarak belirtmiştir. Myford ve Wolfe'a (2003) göre, 1.5 ile 2 arasındaki değerler ölçme modeline katkı sağlamamakta; ancak olumsuz yönde de etkilememektedir. 2'nin üstündeki değerler ise verilerin ölçüm için uygun olmadığını yansıtmaktadır (Sudweeks, Reeveb & Bradshawc, 2004). Tablo 6'daki uyum istatistiklerinin, yukarıda ifade edilen tüm kabul edilebilir sınırlar içerisinde kaldığı göz önüne alındığında model-veri uyumunun sağlandığı söylenebilir.

Tablo 6'nın en alt satırında güvenilirlik katsayısı, ayırma oranı (separation index) ve ki-kare değerleri bulunmaktadır. Tablodaki ki-kare değerinin anlamlı olması ( $\chi^2=453.4$ ,  $sd=184$ ,  $p<0.001$ ), konuların güçlük düzeyi ile ilgili farklı düşüncelere sahip olan öğrencilerin birbirinden anlamlı şekilde ayırt edilebildiğini göstermektedir. Tablo 6'ya göre güvenilirlik katsayısı 0.70 ve ayırma oranı 1.53 olarak hesaplanmıştır. Buradaki güvenilirlik katsayısı Cronbach alfa ya da genellenebilirlik katsayısına eş değer şekilde yorumlanmaktadır (Nakamura, 2002). Bu noktadan hareketle; Walker, Engelhard ve Thompson (2012) Cronbach alfa iç tutarlık katsayısında olduğu gibi Rasch analizinden elde edilen güvenilirlik katsayısı için de 0.70 değerlerinin ölçüt olarak alınmasını önermiştir. Tablodaki güvenilirlik katsayısı bu ölçütü karşılamaktadır. Ayırma oranı için ise 2 ve üzerindeki değerlerin kabul edilebilir olduğu belirtilmektedir (Linacre, 2012). Tablodaki ayırma oranı incelendiğinde bu ölçütün karşılanmadığı anlaşılmaktadır. Güvenirlik katsayısı 0.70 ölçütünü sağlamakta birlikte çok yüksek olmaması ve ayırma oranının 2'nin altında yer alması öğrencilerin çok yüksek güvenilirlikte birbirinden ayırt edilemediği anlamına gelmektedir. Bu durumun öğrencilerin ölçme değerlendirme dersindeki konuların güçlük düzeyi hakkında benzer düşüncelere sahip olmasından kaynaklandığı düşünülmektedir. Daha açık bir anlatımla ayırma oranı ve güvenilirlik değerlerinin çok yüksek çıkmaması, öğrencilerin ölçme değerlendirme dersindeki konuların güçlük düzeylerine ilişkin algıları açısından homojen bir yapıya sahip olmasıyla açıklanabilir. Öğrencilere ait ölçümlerin ardından Tablo 7'de ölçme değerlendirme dersi konularına ilişkin ölçüm raporları incelenmiştir.

Tablo 7.

*Ölçme ve değerlendirme dersi konularına ilişkin ölçüm raporu*

Konular	Ölçüm	Std. Hata	Uyum-içi	Uyum-dışı
1. Temel Kavramlar	-2.73	0.17	1.12	1.20
2. Ölçmede Hata	-1.76	0.15	0.98	0.91
3. Ölçme Araç. Bul. Ger. Nitelikler	-0.54	0.14	0.98	1.11
4. Ölçme Araç ve Yöntemleri	-0.50	0.14	0.85	1.01
5. Test Hazırlama ve Madde Analizi	2.13	0.16	0.98	0.88
6. Test Puanları İstatistikleri	3.42	0.20	1.02	1.07
Ortalama	0.00	0.16	0.99	1.03
Standart Sapma	2.13	0.02	0.08	0.11
Güvenirlik = 0.99	Ayırma Oranı = 13.10	Ki-kare = 871.3	sd = 5	p = 0.00

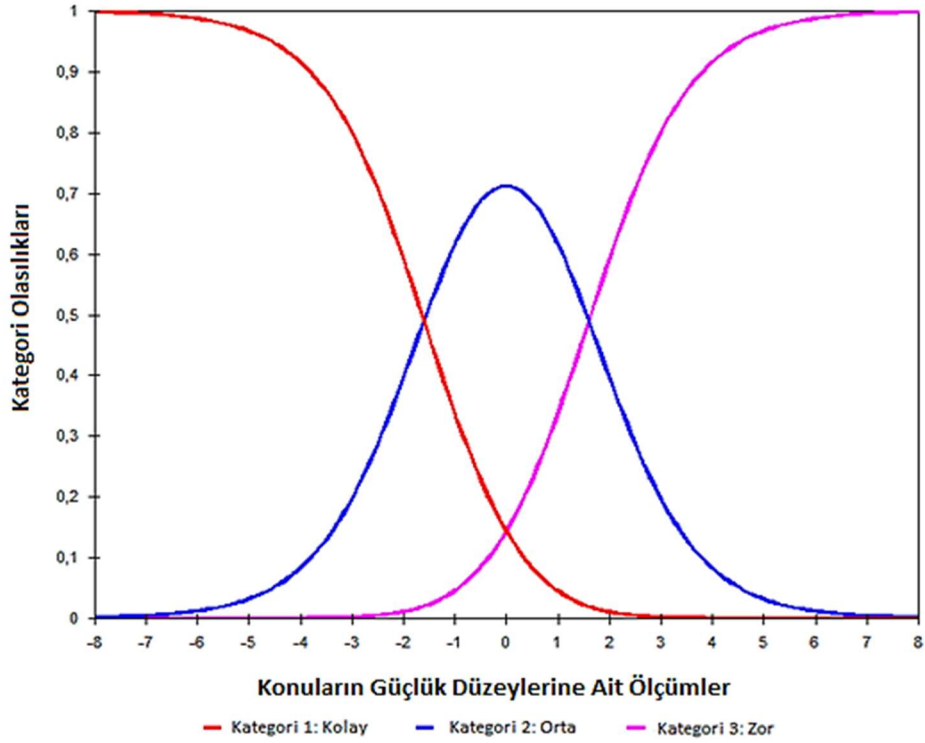
Tablo 7'deki bilgilere göre (Şekil 2'den de görüleceği üzere), 3.42 logit değeri ile altıncı konu öğrencilerin en zor olduğunu düşündükleri konudur. Logit değeri -2.73 olan birinci konu ise öğrenciler tarafından en kolay olarak algılanan konudur. Konular için hesaplanan uyum istatistikleri incelendiğinde hem uyum-içi hem de uyum-dışı istatistiklerinin 0.5 ile 1.5 sınırları içerisinde kaldığı saptanmıştır. Ölçme değerlendirme dersi konuları için rapor edilen ki-kare testinin anlamlı çıkması ( $\chi^2=871.3$ ,  $sd=5$ ,  $p<0.001$ ), güçlük düzeyleri açısından konular arasında anlamlı fark bulunduğunu göstermektedir. Tablo 7'ye bakıldığında konulara ilişkin güvenilirlik katsayısının 0.99 ve ayırma oranının 13.10 olarak bulunduğu görülmektedir. Güvenirlik katsayısı ve ayırma oranı için hesaplanan bu katsayılar, güçlük düzeyleri farklı olan konuların yüksek güvenilirlikte ayırt edilebildiğini yansıtmaktadır. Bu bulgulara ek olarak, konuların güçlük düzeylerine ait sıralamaların konuların derste işleniş sırası ile paralellik gösterdiği saptanmıştır. Derste işleniş sırasına göre ilk konudan son konuya gidildikçe konuların güçlük düzeyi artmaktadır. Tablo 7'nin ortaya koyduğu bir diğer bulgu, güçlük düzeyi birbirine en yakın olan konularla ilgilidir. Üçüncü ve dördüncü konular -0.54 ve -0.50 logit değerleriyle güçlük düzeyleri birbirine en yakın konular olarak belirlenmiştir. Rasch analizi kapsamında incelenen son çıktı; kolay (1), orta (2) ve zor (3) şeklinde kullanılan üçlü dereceleme ait kategori istatistikleridir. Kategori istatistiklerine ilişkin bulgular Tablo 8'de sunulmuştur.

Tablo 8.

*Çalışmada kullanılan üçlü derecelmeye ait kategori istatistikleri*

Kategoriler	Frekans	Yüzde	Yığılmalı Yüzde	Ortalama Ölçüm	Beklenen Ölçüm	Uyum Dışı
1 (kolay)	331	30	30	-2.42	-2.47	1.0
2 (orta)	440	40	70	-0.25	-0.15	1.1
3 (zor)	333	30	100	2.93	2.85	0.9

Ölçme işleminde kullanılan derecelemenin problemsiz bir şekilde çalıştığı söylenmesi için karşılanması gereken bazı şartlar vardır. Bu şartlar; *i*) ölçek kategorilerinin her birinde en az 10 gözlem bulunması ve kategori dağılımlarının düzenli olması, *ii*) ölçek kategorileri ile birlikte ortalama ölçüm değerlerinin de artması ve *iii*) uyum dışı istatistiklerinin 0.5 ile 1.5 kabul edilebilir aralığı içerisinde kalması şeklinde sıralanabilir (Linacre, 2014). Tablo 8 incelendiğinde bu üç şartın da sağlandığı görülmektedir. Dolayısıyla, ölçekte kullanılan derecelemenin etkin çalıştığı ve kullanılan kategori sayısının makul olduğu söylenebilir. Ölçekteki kategori sayısının yeterli olduğunu ve kullanılan derecelemenin sorunsuz çalıştığını ortaya koyan başka bir bulgu da, *dereceleme ölçeğindeki kategorilerin olasılık eğrisidir*. Bu olasılık eğrisi Şekil 3'te gösterilmiştir. Şekil 3'teki yatay eksen konuların güçlük düzeyini yansıtmaktadır. Dikey eksen ise ölçek kategorilerinin kullanım olasılığını temsil etmekte olup 0 ile 1 arasında değişen değerler alabilmektedir.



Şekil 3. Dereceleme ölçeğindeki kategorilerin olasılık eğrisi

Şekil 3 incelendiğinde kolay olduğu düşünülen konularda kategori 1'in, zor olduğu düşünülen konularda ise kategori 3'ün kullanılma olasılığının daha yüksek olduğu görülmektedir. Diğer bir deyişle, eğrinin yatay eksenini boyunca soldan sağa doğru gidildikçe konuların güçlük düzeyi artmakta ve bu artış sırasında kategori 1'in kullanım olasılığı düşerken; kategori 3'ün kullanım olasılığı yükselmektedir. Yine Şekil 3'e göre, kategori 2'nin kullanım olasılığının en yüksek olduğu noktalar orta güçlükteki konuların bulunduğu aralığa düşmektedir. Zor konularda dereceleme ölçeğinin üst ucundaki kategorinin; kolay konularda ise dereceleme ölçeğinin alt ucundaki kategorilerin kullanılma olasılığının daha fazla olması, ölçeğin etkin bir biçimde çalıştığına yönelik bir kanıt olarak değerlendirilebilir.

## TARTIŞMA, SONUÇ VE ÖNERİLER

Bu araştırmada, ikili karşılaştırmalarla ölçekleme tekniği ile Rasch analizinden elde edilen ölçek değerlerinin karşılaştırılması amaçlanmıştır. Bu amaç doğrultusunda öncelikle katılımcılardan ölçme ve değerlendirme dersi kapsamında işlenen altı konuyu ikişerli karşılaştırmaları ve karşılaştırdıkları iki konudan hangisinin kendileri için daha zor olduğunu belirtmeleri istenmiştir. Bu şekilde toplanan veriler üzerinden ikili karşılaştırmalarla ölçekleme çalışması yürütülmüştür. Bunun yanı sıra katılımcıların aynı altı konuyu, güçlük düzeyleri bakımından kolay, orta ve zor şeklinde derecelmeleri istenmiştir. Bu uygulamadan elde edilen veriler ise Rasch modeline göre analiz edilmiştir. Araştırmada, konuların güçlük düzeyleri için ikili karşılaştırmalarla ölçekleme yöntemi ve Rasch analizi sonucunda ulaşılan sıralamaların aynı olduğu görülmüştür. Hem ikili karşılaştırmalarla ölçekleme yönteminde hem de Rasch analizinde ölçme değerlendirme dersinin ilk konusundan son konusuna doğru gidildikçe konuların güçlük düzeylerinin arttığı saptanmıştır. Sıralamanın yanı sıra konuların güçlük düzeylerinin birbirlerine yakınlıkları da her iki yöntemde aynı bulunmuştur. Bir diğer ifadeyle konuların güçlük düzeylerine ilişkin eşit aralıklı ölçek düzeyinde hesaplanan ölçek değerlerinin birbirlerine yakınlığı/uzaklığı da her iki yöntemde benzer çıkmıştır. Örneğin, her iki yöntemde üçüncü ve dördüncü konular güçlük düzeyleri bakımından birbirine en yakın konular olarak belirlenmiştir. Bu sonuçlar, ikili karşılaştırmalarla ölçekleme yönteminde ve Rasch analizinde hesaplanan ölçek değerlerinin tutarlı olduğunu ortaya koymaktadır.

Çalışmadan elde edilen bulgular, konu ile ilgili alanyazında yer alan kuramsal bilgiler ile desteklenmektedir. Andrich (1978b) ile Jansen'in (1984) Thurstone'un beşinci hal denklemi ile Rasch modelini kuramsal bir çerçevede matematiksel olarak karşılaştırdıkları çalışmalarda, iki yöntemde hesaplanan parametrelerin aynı olacağı ifade edilmiştir. Araştırmada ulaşılan bulgular Wainer vd.'nin (1978) gerçek veriler üzerinden ampirik olarak yürüttüğü çalışmanın sonuçlarıyla da örtüşmektedir. Wainer vd. (1978) tarafından yapılan çalışmada, Thurstone'un beşinci hal denklemi ile Rasch analizinde hesaplanan ölçek değerlerinin birbirine çok yakın olduğu tespit edilmiştir.

Araştırma sonuçlarından hareketle, sıralama ölçeği ile toplanan verileri eşit aralık düzeyine getirmek için hem ikili karşılaştırmalarla ölçekleme tekniğinin hem de Rasch analizinin kullanılabileceği ve bu iki yöntem ile elde edilecek ölçek değerlerinin tutarlı olacağı söylenebilir. Her iki yöntemde uyarıcılara ilişkin ölçek değerleri ile ilgili yapılacak yorumlar benzer olacağından tercih edilecek yönteme araştırmacı kendisi karar verebilir. Ancak bu karar sırasında yöntemlerin birbirlerine göre üstünlük ve zayıflıkları göz önünde bulundurulmalıdır. İkili karşılaştırmalarla ölçekleme çalışmasında uyarıcılar gözlemcilere ikişerli olarak sunulmakta ve buna bağlı olarak özellikle uyarıcı sayısının fazla olduğu durumlarda katılımcılara düşen iş yükü artmaktadır. Maddeler ikili karşılaştırmalar yerine dereceleme biçiminde katılımcılara sunulduğunda ise veri toplama işlemi daha kısa sürmekte ve ölçme aracının nasıl doldurulması gerektiği katılımcılar tarafından daha kolay anlaşılabilir. Dolayısıyla ikili karşılaştırmalarla ölçekleme yöntemi ile kıyaslandığında Rasch modeline ilişkin verilerin toplanması daha rahat olabilmektedir. Yöntemlerin veri toplama sürecindeki bu farklılık, ikili karşılaştırmalarla ölçeklemeye göre Rasch analizinin avantajlı bir yönü olarak görülebilir. Rasch analizinde, uyarıcılara ilişkin ölçek değerleriyle birlikte çalışmaya konu olan yüzeylere ait detaylı ölçüm raporlarına ulaşılabilmesi, bu yöntemin ikili karşılaştırmalarla ölçekleme çalışmasına göre bir başka üstün yönüdür. Son olarak, ikili karşılaştırmalarla ölçekleme çalışmasına ilişkin analizler Microsoft Office Excel aracılığıyla gerçekleştirilebilmektedir. Diğer taraftan Rasch analizi için kullanılabilecek bilgisayar programlarının büyük çoğunluğuna (örneğin bu araştırmada kullanılan FACETS programı) ancak belirli bir ücret karşılığında erişilebilmektedir. Bundan dolayı, herhangi bir uyarıcı setine ilişkin eşit aralık düzeyinde ölçek değerlerine ulaşmak isteyen fakat Rasch analizini gerçekleştirilebileceği lisanslı bir programa sahip olmayan araştırmacılar için ikili karşılaştırmalarla ölçekleme yönteminin daha uygun bir tercih olabileceğini ifade etmek mümkündür. Kullanılacak yöntem hakkında bir tercih yaparken, bu farklılıkların göz önünde bulundurulması önemlidir.

## KAYNAKÇA/REFERENCES

- Akın Ankan, Ç., Uğurlu, S., & Atar, B. (2016). MIMIC, SIBTEST, lojistik regresyon ve Mantel-Haenszel yöntemleriyle gerçekleştirilen DMF ve yanlılık çalışması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 31(1), 34-52. <http://dx.doi.org/10.16986/HUJE.2015014226>
- Andrich, D. (1978a). A rating formulation for ordered response categories. *Psychometrika*, 43(9), 561-573. <http://dx.doi.org/10.1007/BF02293814>
- Andrich, D. (1978b). Relationships between the Thurstone and Rasch approaches to item scaling. *Applied Psychological Measurement*, 2(3), 449-460. <http://dx.doi.org/10.1177/014662167800200319>
- Anıl, D., Ertuna, L., & Uysal, İ. (2015). Türkiye'deki araştırma görevlilerinin mesleki sorunlarının ikili karşılaştırma yoluyla ölçeklenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 6(2), 279-292. <http://dx.doi.org/10.21031/epod.14618>
- Anıl, D., & Güler, N. (2006). İkili karşılaştırmalar yöntemi ile ölçekleme çalışmasına bir örnek. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 30, 30-36. <http://www.efdergi.hacettepe.edu.tr/yonetim/icerik/makaleler/722-published.pdf> adresinden edinilmiştir.
- Arık, R.S., & Kutlu, Ö. (2013). Öğretmenlerin ölçme ve değerlendirme alanı yeterliklerinin yargıcı kararlarına dayalı ölçeklenmesi. *Eğitim Bilimleri Araştırmaları Dergisi*, 3(2), 163-196. <http://dx.doi.org/10.12973/jesr.2013.3210a>
- Atılgan, H. (2004). *Genellenebilirlik kuramı ve çok değişkenlik kaynaklı Rasch modelinin karşılaştırılmasına ilişkin bir araştırma*. Yayınlanmamış Doktora Tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Atılgan, H. (Ed.). (2013). *Eğitimde ölçme ve değerlendirme*. Ankara: Anı.
- Bahar, M., Nartgün, Z., Durmuş, S., & Bıçak, B. (2010). *Geleneksel-tamamlayıcı ölçme ve değerlendirme teknikleri*. Ankara: Pegem Akademi.
- Baker, F.B. (2001). *The basics of item response theory*. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD.
- Başol, G. (2013). *Eğitimde ölçme ve değerlendirme*. Ankara: Pegem Akademi.
- Baykul, Y. (1996). *İstatistik metotlar ve uygulamalar*. Ankara: Lazer Ofset.
- Baykul, Y. (2010). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Ankara: Pegem Akademi.
- Brentari, E., & Golia, S. (2008). Measuring job satisfaction in the social services sector with the Rasch model. *Journal of Applied Measurement*, 9(1), 45-56. Retrieved from [https://www.researchgate.net/publication/5668940\\_Measuring\\_job\\_satisfaction\\_in\\_the\\_social\\_services\\_sector\\_with\\_the\\_Rasch\\_model](https://www.researchgate.net/publication/5668940_Measuring_job_satisfaction_in_the_social_services_sector_with_the_Rasch_model).
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Florida: Holt, Rinehart and Winston.
- Çıkrıkçı Demirtaşlı, N. (Ed.). (2012). *Eğitimde ölçme ve değerlendirme*. Ankara: Öz Baran Ofset.
- DeMars, C. (2010). *Item response theory*. Oxford, UK: Oxford University.
- Doğan, N. (2002). *Klasik test teorisi ve örtük özellikler kuramının örneklemeler bağlamında karşılaştırılması*. Yayınlanmamış Doktora Tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Doğan, N., & Öğretmen, T. (2008). Değişen madde fonksiyonunu belirlemede Mantel Haenszel, ki-kare ve lojistik regresyon tekniklerinin karşılaştırılması. *Eğitim ve Bilim*, 35(156), 3-16. <http://egitimvebilim.ted.org.tr/index.php/EB/article/viewFile/679/130> adresinden edinilmiştir.
- 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>
- Erkuş, A. (2012). *Psikolojide ölçme ve ölçek geliştirme-I*. Ankara: Pegem Akademi.
- Ertoprak, D.G. (2015). Üniversite tercihlerinde etkili olabilecek özellikleri belirlemeye yönelik bir ölçekleme çalışması. *Amasya Üniversitesi Eğitim Fakültesi Dergisi*, 4(2), 153-174. <http://dx.doi.org/10.17539/aej.96310>
- Gömlüksiz, M., & Erkan, S. (Ed.). (2010). *Eğitimde ölçme ve değerlendirme*. Ankara: Nobel.
- Güler, N. (2012). *Eğitimde ölçme ve değerlendirme*. Ankara: Pegem Akademi.
- Güler, N., & Anıl, D. (2009). Scaling through pair-wise comparison method in required characteristics of students applying for post graduate programs. *International Journal of Human Sciences*, 6(1), 627-639. Retrieved from <https://j-humansciences.com/ojs/index.php/IJHS/article/download/673/365>.
- Hall, R. (2008). *Applied social research: Planning, designing and conducting real-world research*. South Yarra, Australia: Palgrave Macmillan.
- Jackson, S. (2008). *Research methods: A modular approach*. Belmont, CA: Thompson Wadsworth.



- Jansen, P.G.W. (1984). Relationships between the Thurstone, Coombs, and Rasch approaches to item scaling. *Applied Psychological Measurement*, 8(4), 373-383. <http://dx.doi.org/10.1177/014662168400800401>
- Johnson, B., & Christensen, L. (2014). *Educational research: Quantitative, qualitative, and mixed approaches*. Thousand Oaks, CA: Sage.
- Kan, A. (2008a). Yargıcı kararlarına dayalı ölçekleme yöntemlerinin karşılaştırılması üzerine ampirik bir çalışma. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 35, 186-194. <http://www.efdergi.hacettepe.edu.tr/yonetim/icerik/makaleler/552-published.pdf> adresinden edinilmiştir.
- Kan, A. (2008b). Psikolojik değişkenleri ölçmek için kullanılan ölçekleme yaklaşımları üzerine bir karşılaştırma. *Eğitimde Kuram ve Uygulama*, 4(1), 2-18. <http://eku.comu.edu.tr/article/download/1044000044/1044000127> adresinden edinilmiştir.
- Köse, İ.A., & Öztumur, B. (2014). Kayıp veri ele alma yöntemlerinin t-testi ve anova parametreleri üzerine etkisinin incelenmesi. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 14(1), 400-412. <http://dx.doi.org/10.17240/aibuefd.2014.14.1-5000091519>
- Linacre, J.M. (1989). *Many-facet Rasch measurement*. Chicago: MESA.
- Linacre, J.M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3(1), 85-106. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.424.2811&rep=rep1&type=pdf>.
- Linacre, J.M. (2012). *Many-facet Rasch measurement: Facets tutorial*. Retrieved from <http://www.winsteps.com/a/ftutorial2.pdf>.
- Linacre, J.M. (2014). *A user's guide to FACETS Rasch-model computer programs*. Retrieved from <http://www.winsteps.com/a/facets-manual.pdf>.
- Lunz, M.E., & Stahl, J.A. (1990, April). *Severity of grading across time periods*. Paper presented at the Annual Meeting of the American Educational Research Association, Boston. Retrieved from <http://files.eric.ed.gov/fulltext/ED317602.pdf>.
- Lunz, M.E., Wright, B.D., & Linacre, J.M. (1990). Measuring the impact of judge severity on examination scores. *Applied Measurement in Education*, 3(4), 331-345. [http://dx.doi.org/10.1207/s15324818ame0304\\_3](http://dx.doi.org/10.1207/s15324818ame0304_3)
- Martella, R.C., Nelson, J.R., Morgan, R.L., & Marchand Martella, N.E. (2013). *Understanding and interpreting educational research*. New York: The Guilford.
- Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149-174. <http://dx.doi.org/10.1007/BF02296272>
- Myford, C.M., & Wolfe, E.W. (2003). Detecting and measuring rater effects using many-facet Rasch measurement: Part I. *Journal of Applied Measurement*, 4(4), 386-422. Retrieved from [https://www.researchgate.net/publication/9069043\\_Detecting\\_and\\_Measuring\\_Rater\\_Effects\\_Using\\_Many-Facet\\_Rasch\\_Measurement\\_Part\\_I](https://www.researchgate.net/publication/9069043_Detecting_and_Measuring_Rater_Effects_Using_Many-Facet_Rasch_Measurement_Part_I).
- Nakamura, N. (2002). Teacher assessment and peer assessment in practice. *Educational Studies*, 44, 143-204-215. Retrieved from [https://ia800200.us.archive.org/4/items/ERIC\\_ED464483/ERIC\\_ED464483.pdf](https://ia800200.us.archive.org/4/items/ERIC_ED464483/ERIC_ED464483.pdf).
- O'Neill, T.R., & Lunz, M.E. (1996, April). *Examining the invariance of rater and project calibrations using a multi-facet Rasch model*. Paper Presented at the Annual Meeting of the American Educational Research Association, New York. Retrieved from <http://files.eric.ed.gov/fulltext/ED398284.pdf>.
- Polat, B., & Göksel, H.G. (2014). Öğretmen adaylarının sosyal aktivite tercihlerinin ikili karşılaştırmalı ölçekleme yöntemiyle belirlenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 5(1), 88-100. <http://epoddergipark.gov.tr/download/article-file/65953> adresinden edinilmiştir.
- Powell, R.R. (1997). *Basic research methods for librarians*. Greenwich, CT: Ablex.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Danish Institute for Educational Research.
- Schaefer, E. (2008). Rater bias patterns in an EFL writing assessment. *Language Testing*, 25(4), 465-493. <http://dx.doi.org/10.1177/0265532208094273>
- Sudweeks, R.R., Reeve, S., & Bradshaw, W.S. (2005). A comparison of generalizability theory and many-facet Rasch measurement in an analysis of college sophomore writing. *Assessing Writing*, 9(3), 239-261. <http://dx.doi.org/10.1016/j.asw.2004.11.001>
- Şahin Kürşad & Nartgün, (2015). Kayıp veri sorununun çözümünde kullanılan farklı yöntemlerin ölçeklerin geçerlik ve güvenilirliği bağlamında karşılaştırılması. *Eğitimde ve psikolojide ölçme ve değerlendirme dergisi*, 6(2), 25-267. <http://dx.doi.org/10.21031/epod.95917>
- Tan, Ş. (2015). *Öğretimde ölçme ve değerlendirme KPSS el kitabı*. Ankara: Pegem Akademi.
- Tekindal, S. (2009). *Duyuşsal özelliklerin ölçülmesi için araç oluşturma*. Ankara: Pegem Akademi.

- Thurstone, L.L. (1927). A law of comparative judgment. *Psychological Review*, 34(4), 273-286.  
<http://dx.doi.org/10.1037/h0070288>
- Torgerson, W.S. (1958). *Theory and methods of scaling*. Newyork: John Wiley & Sons Inc.
- Turgut, M.F., & Baykul, Y. (1992). *Ölçekleme teknikleri*. Ankara: ÖSYM.
- Wainer, H., Fairbank, D.T., & Hough, R.L. (1978). Predicting the impact of simple and compound life change events. *Applied Psychological Measurement*, 2(3), 311-320.  
<http://dx.doi.org/10.1177/014662167800200301>
- Walker, E.R., Engelhard, G., & Thompson, N.J. (2012). Using Rasch measurement theory to assess three depression scales among adults with epilepsy. *Seizure*, 21(6), 437-443.  
<http://dx.doi.org/10.1016/j.seizure.2012.04.009>
- Wright, B.D. & Linacre, J.M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8, 370. Retrieved from <http://www.rasch.org/rmt/rmt83b.htm>.
- Wright, B.D. & Masters, G. (1982). *Rating scale analysis: Rasch measurement*. Chicago: MESA.

### **İletişim/Correspondence**

Doç. Dr. Neşe GÜLER  
[gnguler@gmail.com](mailto:gnguler@gmail.com)

Yrd. Doç. Dr. Mustafa İLHAN  
[mustafailhan21@gmail.com](mailto:mustafailhan21@gmail.com)

Yrd. Doç. Dr. Gülşen TAŞDELEN TEKER  
[gulsentasdelen@gmail.com](mailto:gulsentasdelen@gmail.com)