



## Sınav Sorularının Geliştirilmesinde Rasch Analizi Kullanımı

**Zeynep ÖZTÜRK**

Dr. Öğr. Üyesi, Artvin Çoruh Üniversitesi

zozturk@artvin.edu.tr

<https://orcid.org/0000-0003-0475-1357>

Makale Başvuru Tarihi : 05.11.2024

Makale Kabul Tarihi : 16.12.2024

Makale Yayın Tarihi : 31.12.2024

Makale Türü : Araştırma Makalesi

DOI: 10.5281/zenodo.14583363

### Özet

#### Anahtar Kelimeler:

Rasch Analizi,  
Eğitimde Ölçme  
ve Değerlendirme,  
Test Geliştirme

Eğitimciler öğrencileri, öğrenci becerilerini, yeterliliklerini sınav, test ve quiz gibi değerlendirme süreçlerini kullanarak ölçmek isterler. Bu çalışmanın amacı, eğitim süreçlerinde sınavların ölçme geçerliliğini ve güvenilirliğini artırmak için Rasch analizi yönteminin uygulamalı bir örneğini sunmak ve sınav sorularının zorluk düzeylerini ve ayırt edicilik özelliklerini belirlemedeki rolünü ortaya koymaktır. Bu çalışmada kullanılan ölçme aracı, Türkiye'deki bir üniversitede yabancı dil muafiyet sınavına giren 650 üniversite 1. Sınıf öğrencisinden oluşmaktadır. Testin özellikleri Rasch modeli kullanılarak incelenmiştir. 20 maddeden oluşan nihai testin güvenilirlik düzeyi 0,74 olarak hesaplanmıştır. Test sorularının modele uygunluğu için tek boyutluluk, değişmezlik, Rasch modele uyum, model güvenilirliği testleri uygulanmıştır. Verinin modele uygunluğu için varsayımlar ( tek boyutluluk, değişmezlik, model uyumu, güvenilirlik) sınanmış ve 6 soru analizden çıkarılmıştır. 14 soru için öğrencilerin yanıtlama düzeyleri ve soruların zorluk parametreleri elde edilmiştir. En zor soru s12 ve en kolay soru ise, s5 olarak bulunmuştur. Soruların %7,14'ü çok kolay ve çok zor, %14,2'si kolay, %50'si orta ve %21,4'ü zor sorulardan oluşmaktadır. Yetenek değerleri, genel olarak 0.540 ile 0.903 arasında değişmekte ve testin maddelerinin ayırt edicilik açısından makul düzeyde olduğunu göstermektedir. Wright Harita, test sorularının çoğunlukla katılımcıların yetenek seviyelerine uygun olduğunu, ancak aşırı yüksek veya düşük yetenek düzeyine sahip bireyler için uygun maddelerin eksik olduğunu göstermektedir. Testin daha geniş bir yetenek aralığını kapsamaları için hem daha kolay hem de daha zor sorular eklenmelidir.

### Using Rasch Analysis in the Development of Exam Questions

#### Keywords:

Rasch Analysis,  
Educational  
Measurement, Test  
Development

#### Abstract

Educators want to measure students, student skills and competencies by using assessment processes such as exams, tests and quizzes. The aim of this study is to present an applied example of the Rasch analysis method to increase the measurement validity and reliability of exams in educational processes and to show its role in determining the difficulty levels and discriminative features of exam questions. The measurement tool used in this study consists of 650 first-year university students who took a foreign language exemption exam at a university in Turkey. The properties of the test were analysed using the Rasch model. The reliability of the final 20-item test was calculated to be 0.74. Unidimensionality, invariance, Rasch model fit and model reliability tests were applied to the model fit of the test items. Assumptions (unidimensionality, invariance, model fit, reliability) were tested for model fit of the data and 6 questions were removed from the analysis. For 14 questions, student response levels and question difficulty parameters were obtained. The most difficult question was found to be p12 and the easiest question was found to be p5. 7.14% of the questions were very easy and very difficult, 14.2% were easy, 50% were medium and 21.4% were difficult. The ability values generally range between 0.540 and 0.903, indicating that the items of the test have a reasonable level of discrimination. The Wright Map shows that the test questions are mostly appropriate for the participants' ability levels; however, there is a lack of suitable items for individuals with extremely high or low ability levels. To cover a wider range of abilities, both easier and more difficult questions should be added to the test.

## 1. GİRİŞ

Tüm eğitim kurumlarının ortak bir özelliği, başarılı ve öğrendiklerini toplumlarına ve ülkelerine yardımcı olmak için uygulayabilen öğrenciler yetiştirmektir. Öğrenci başarısı, tüm eğitimciler, öğrenciler, üniversiteler ve diğer ilgili taraflar için önemli bir husustur. Öğrencilerin bir konuyu anlamaları, öğrenmenin en önemli bileşenidir. Eğitimciler, öğrencinin bilişsel düzeyini orantılı bir şekilde değerlendirmek için bir ölçme aracı kullanır. İyi ölçme aracı (sınav), öğrencilerin performansını ölçmek için hedef nitelikleri değerlendirebilecek uygun maddelerin seçimiyle olur. Rasch analizi kullanımı, ölçme araçlarının hedef kitleye uygunluğunu, temel ölçüm gerekliliklerine uygun olup olmadığını, her bir sorunun doğası hakkında bilgi elde etmeyi, öğrenci becerilerine ilişkin güvenilir kanıtlar sağlayan, bir testin geliştirilmesine olanak sağlar (Rasch,1960). Rasch analizi özellikle eğitim ve psikolojik alan testleri analizinde sıkça kullanılmaktadır. Bu yöntem, bir bireyin bir teste verdiği yanıtların, bireyin yetenek düzeyi ile test maddelerinin zorluk dereceleri arasındaki farkın bir fonksiyonu olarak modellenmesine olanak tanır (Hambleton ve Swamington, 1985:13). Rasch modeli, ölçme sürecinde yanlılığı azaltarak test maddelerinin daha adil bir şekilde değerlendirilmesini sağlar ve test katılımcılarının bilgi düzeylerini daha nesnel bir şekilde analiz etme imkânı sunar. Bu özellikleri sayesinde Rasch analizi, sınav sorularının adil ve tutarlı bir şekilde geliştirilmesi sürecinde kullanılan bir araçtır (Dhyaaldian vd., 2022:100). Eğitimcilerin özellikle çoktan seçmeli soruları yazma ve her bir sorunun zorluk seviyesini belirleme süreci de dâhil olmak üzere test planlaması için saatler harcamaları gerekir. Bu, belki de öğretmenler tarafından geliştirilen çoğu test maddesinin neden yüksek ve düşük yetenek arasında ayırım yapamadığının ve maddelerin amaçlanan kullanımlarına göre işlev görmediğinin açıklanmasının olası nedenlerinden biri olarak hizmet edebilir (Richichi, 1996:2210). Modern ölçüm teorisindeki Rasch modeli, bu tür test sorularının kalitesini analiz etmek için yeni bir yol sunmakta ve ayırt edici soruların analize uygulanabilirliğini sağlamaktadır. Sınav sorularının kalite standartlarını sağlamak, değerlendirme sürecinin güvenilirliğini artırmak ve yanlılıklardan arındırılmış bir ölçüm yapmak için bilimsel yöntemler ve analizler kullanılmalıdır.

Literatürde, Draugalis vd. (2004) öğrenci ve madde performansını değerlendirmek ve müfredatın güçlü ve zayıf yönlerini 65 maddelik çoktan seçmeli bir sınav kullanarak değerlendirmek için Rasch Modelini uyguladı. Başka bir çalışmada Azrilah vd. (2008) ölçüm araçlarının yapısını doğrulamak için Rasch Modelini kullandı. Rasch ölçümünün bir tutumun uzman düzeyindeki mutabakatını anlamada daha iyi bir keşif derinliği sağladığı bulundu. Zulkifli vd. (2010) Malezya Kebangsaan Üniversitesi'ndeki doğrusal cebir dersine ilişkin çok amaçlı sorularda madde yanıtının 0-1 puanlamasını kullanan ikili bir Rasch modeli uyguladı. Zulkifli vd. (2012) Rasch Modelinin hem öğrenci yeteneğini hem de soru geçerliliğini ölçmek için kullanılmaya uygun olduğu sonucuna vardı. Rasch modelini kullanarak belirli bir dersin belirtilen hedeflerine ilişkin sınav sorularının güvenilirliğini ölçmeye odaklandı. Bu çalışmanın arkasındaki fikir, öğrencilerin başarısını uygun soru setleriyle artırmaktı. Ayrıca, Oliveira vd. (2012) Rasch modelin cloze maddelerin analizine uygulanmasını araştırmıştır. Boone (2016) ise, Rasch analizini test geliştirmeye rehberlik edebilir, puan anlamlarını açıklayabilir ve farklı test formlarının oluşturulmasına olanak sağlayabilmesi üzerine çıkarımda bulunmuştur. Chen ve Zhou (2018) Rasch modelini Çin'deki Ulusal Matrikülasyon İngilizce Testindeki (NMET) cloze testi incelemek için kullanmıştır. Gan vd. (2018)'de Çin'deki 2015 ve 2016 İngilizce Ana Dilleri Testi-Band 4 (TEM4) testini incelemiştir. Baghaei ve Ravand (2019) İran'daki Ulusal Üniversite Giriş Sınavını bu model ile araştırmış; Mokshein vd. (2019) Malezya'daki devlet okulları ve özel okullardaki 6. sınıf öğrencilerine Milli Eğitim Bakanlığı tarafından uygulanan İlkokul Başarı Testi'nin İngilizce maddelerinin kalite analizi için kullanmıştır. Dhyaaldian vd. (2022) Irak'taki bir üniversitede okuduğunu anlama dersinin ara sınavında cloze testinde denemiştir. Türkiye'de Rasch temelli ölçme aracı geliştirme tezleri incelendiğinde, tek boyutluluk, yerel bağımsızlık ve farklı madde işleyişi gibi temel model varsayımlarının ve analizlerinin raporlanmasında tutarsızlıklar bulunmuştur (Tunç, 2023:265). Rasch tekniklerinin uygulanmasında daha sistematik bir yaklaşıma olan ihtiyaç vardır.

Bu çalışmanın temel gerekçesi, literatürde sınav sorularının geliştirilmesinde Rasch analizi kullanımına yönelik sınırlı sayıda çalışmanın bulunmasıdır. Mevcut literatürde bu yöntemin pratik uygulamalar ile eğitim süreçlerine somut katkılarının yeterince incelenmediği görülmektedir. Bu eksiklik, sınav sorularının geliştirilmesinde daha bilimsel ve sistematik yaklaşımlara duyulan ihtiyacı ortaya koymaktadır. Ayrıca, mevcut eğitim sistemlerinde sınavların ölçme geçerliliğini artırmaya yönelik uygulamalı çalışmaların yetersizliği, bu çalışmanın yapılmasını gerektiren bir diğer önemli faktördür.

Bu çalışmada, sınav sorularının geliştirilmesinde Rasch analizinin etkili bir araç olarak nasıl kullanılabileceği incelenecektir. Çalışmanın temel amaçları şunlardır:

1. Rasch analizinin sınav sorularının zorluk düzeylerini ve ayırt edicilik özelliklerini belirlemedeki rolünü ortaya koymak.
2. Eğitim süreçlerinde sınavların ölçme geçerliliğini ve güvenilirliğini artırmak için Rasch analizi yönteminin uygulamalı bir örneğini sunmak.

Bu bağlamda, çalışmanın bulguları, eğitim politikaları ve sınav geliştirme süreçlerine yönelik öneriler sunarak eğitimin kalitesinin artırılmasına katkıda bulunmayı hedeflemektedir.

## 2.ARAŞTIRMA YÖNTEMİ

### 2.1.Örnekleme ve Ölçme Aracı

Bu çalışmada, üniversite 1. sınıfa başlayacak öğrencilerin İngilizce dersi muafiyet sınavına ilişkin yanıtları analiz edilmiştir. Muafiyet sınavı, öğrencilerin İngilizce seviyelerini değerlendirmek amacıyla hazırlanmış, toplamda 20 sorudan oluşan çoktan seçmeli bir testtir. Sınavda her bir soruya verilen yanıtlar, doğru veya yanlış şeklinde değerlendirilmiş ve yanıtlar doğruluğuna göre 1 (Doğru) ve 0 (Yanlış) olarak kodlanmıştır. Böylece, sınav sonuçlarından ikili (dichotomous) bir veri seti elde edilmiştir. Araştırma, 650 öğrencinin her bir soruya verdiği yanıtların analizini içermektedir ve öğrencilerin İngilizce bilgi düzeylerinin belirlenmesi amacıyla test sorularına detaylı bir değerlendirme süreci gerçekleştirilmiştir. Bu kapsamda, elde edilen veriler Madde Tepki Teorisi modellerinden biri olan Rasch model analizi ile incelenmiş ve sınav sorularının özellikleri ile öğrencilerin yetenek düzeyleri arasındaki ilişki ortaya koyulmaya çalışılmıştır.

### Rasch Analizi

Rasch analizi, test sorularının parametrelerini ve sorular için olası yanıtlar doğrultusunda ölçülmesi amaçlanan özellikleri tahmin eden tek parametrelili bir Madde Tepki Modeli (IRT) dir. Yetenek ve parametre tahminleri, testin uygulandığı örneklemden bağımsızdır. Rasch analizinde bilgi, kişinin yeteneği ile soru zorluğu arasındaki farkın bir fonksiyonudur.

Rasch analizi, Bir parametrelili lojistik modele dayalı IRT modellerinden biridir. Bir parametrelili lojistik model için (1) denklemi

$$P_i(\theta) = \frac{e^{(\theta-b_i)}}{1+e^{(\theta-b_i)}}, \quad i = 1,2, \dots, n \quad (1)$$

olarak tanımlanır. Burada,  $P_i(\theta)$ ,  $\theta$  yeteneği ile rasgele seçilmiş bireyin  $i$ -yinci maddeye doğru bir şekilde cevap verme olasılığını,  $b_i$ ,  $i$  – yinci madde güçlük parametresini,  $n$ , testteki madde sayısını,  $e$  değeri, 2,718 olan bir sayı ve  $P_i(\theta)$  ,yetenek ölçümü ile ilgili 0 ve 1 arasında değer alan fonksiyondur. Bir madde için  $b_i$  parametresi, doğru bir yanıtın olasılığı 0,5 olan yetenek ölçüsündeki noktadır (Hambleton ve Swamington, 1985:12-14; Embretson ve Reise, 2000:20) Bu çalışmada, Rasch modeli kullanılarak Jamovi programı ile analizler gerçekleştirilmiştir. Rasch modeli, tek boyutluluk, yerel bağımsızlık, test ve madde istatistiklerini elde etmek ve testin geçerliliğini ve güvenilirliğini değerlendirmek, verilerin modele uyumu, madde zorluğu ve kişi yeteneği gibi analizler yapılmalıdır.

### 2.2.Analiz Süreci

Bu çalışmada gerçekleştirilen analizler iki temel çerçevede ele alınmıştır: varsayımların sınanması ve parametre tahminlerinin elde edilmesi. Araştırmanın analitik kısmında, Rasch analizi kullanılarak veriler derinlemesine incelenmiştir. Bu analiz, R programlama dilinin sağladığı güçlü istatistiksel altyapıyı, kullanıcı dostu bir arayüz ile birleştiren açık kaynaklı bir istatistiksel analiz yazılımı olan *Jamovi* üzerinde gerçekleştirilmiştir. Analiz sürecinde, *SnowRMM* ve *SnowIRT* modüllerinden yararlanılmıştır. Özellikle, *SnowIRT* modülü, R yazılımında yer alan ‘erm’ paketini temel almakta ve bu sayede madde uyum istatistikleri, Wright haritası, model uyumu, tek boyutluluk, değişmezlik ve Anderson’un Olabilirlik testi gibi kapsamlı analizleri gerçekleştirilebilme imkanı sunmaktadır (Mair vd., 2023:50).

Araştırmada, her bir sorunun testin temel yapısına ne ölçüde uygun olduğunu değerlendirmek amacıyla çeşitli varsayım testleri uygulanmıştır. Bu doğrultuda, sorulara ilişkin madde uyum istatistikleri, nokta ölçüm korelasyonu analizi, model güvenilirliği ve değişmezlik analizleri gerçekleştirilmiştir. Testi oluşturan soruların aynı temel özelliği ölçüp ölçmediğini değerlendirmek amacıyla tek boyutluluk analizleri Martin-Löf olabilirlik oranı testi ile yapılmıştır. Ayrıca Faktör Analizi de kullanılmıştır, böylece soruların aynı temel yapıyı ölçüp ölçmediği sistematik bir şekilde analiz edilmiştir. Rasch modelinin veri setine uygunluğunu

belirlemek amacıyla Anderson Olabilirlik testi ile soruların modele uyumu detaylı bir şekilde incelenmiştir. Çalışmanın bir diğer önemli boyutu, soruların test katılımcılarına uygunluğunun ve soruların zorluk düzeylerinin değerlendirilmesidir. Bu amaçla, güvenilirlik katsayısı ve ayırma indeksi gibi metrikler kullanılmıştır. Ayrıca, Wald testi ile öğrencilerin yeteneklerinin dağılımı (değişmezlik) ve Wright haritası üzerinden soruların güçlük seviyeleri analiz edilmiştir. Özellikle, Wright haritası, testteki soruların zorluk düzeylerinin, katılımcıların yetenek seviyeleri ile olan ilişkisini görselleştirmede etkili bir araç olarak kullanılmıştır. Tüm bu analizler, testin genel yapısının, soruların ve katılımcıların performansına dayalı olarak daha kapsamlı bir değerlendirme yapılmasını sağlamıştır.

### 3.BULGULAR

#### 3.1.Tek Boyutluluk

Sadece bir yeteneğin bir testteki soruların bir kümesi ile ölçülmesine tek boyutluluk denir. Yani, kişinin testteki performansını belirleyen faktörün testin ölçmeye çalıştığı faktör olmasıdır. Bireyin performansını etkileyen birden fazla faktör olduğunda tek boyutluluk sağlanamaz. Rasch analizi sadece tek faktörlü testlerde geçerli sonuçlar verebilir. Tek boyutluluk analizi yapı geçerliliğinin belirlenmesine yönelik değerlendirilen bir bileşendir. Rasch modeline göre bir ölçme aracının tek bir özelliğin ölçümüne katkıda bulunması beklendiğinden tek boyutlu bir yapıya sahip olması gerekmektedir (Bond vd., 2015; J. M. Linacre, 2023). Tek boyutluluk faktör analizi ve aynı zamanda Martin-Löf (MLOef) Olabilirlik Oranı (LR) testiyle de yapılabilir. Faktör analizinde faktör çıkarım yöntemi olarak “temel eksen faktör” ve döndürme yöntemi olarak “oblimin” yöntemi kullanıldı. Özdeğeri 1’den büyük ve faktör yükü 0,30’ dan büyük olma şartlarıyla yapıldı. Tablo 1’de faktör analizi sonuçları verilmiştir.

**Tablo 1. Faktör Boyutu ve Fakör Yük Değerleri**

Faktör Yükleri	Soru	s6	s7	s10	s9	s8	s15	s11	s13	s3	s16
	<b>Faktör</b>		0,470	0,429	0,425	0,404	0,384	0,381	0,366	0,365	0,355
Faktör Yükleri	Soru	s12	s14	s4	s1	s5	s2	s19	s20	s17	s18
	<b>Faktör</b>		0,343	0,340	0,334	0,331	0,325	0,324	0,321	0,301	

Tablo 1’e göre, tek boyut ortaya çıktı ancak, s17 ve s18 sorularının madde faktör yükleri 0,30’ un altında bulunduğundan testi temsil etme gücü düşük olduğundan analizden çıkarılmıştır. Ayrıca, s17 ve s18 analizden çıkarıldıktan sonra tek boyutluluk, Martin-Löf (MLOef) Olabilirlik Oranı (LR) testiyle de kontrol edilmiştir. Tablo 2’de Martin-Löf Olabilirlik Oran testi sonuçları görülür.

**Tablo 2. Martin-Loef Test – Medyan**

	Değer	Serbestlik Derecesi	p
<b>Olabilirlik Oranı</b>	54,6	48	0,239

MLOef testi, medyan madde ham puanlarına göre maddeleri iki alt gruba ayırır ve iki alt grubun homojen olup olmadığını test eder. Tablo 2’den sonuçlar şu şekildedir: LR değeri=54,6,df = 48,  $p = 0,239$  ( $p > 0,05$ ) olduğundan iki alt öğrenci grubuna dayalı madde güçlük ölçümleri istatistiksel olarak aynıdır; bu da madde ölçümlerinin tek boyutlu olduğunu gösterir.

#### 3.2.Değişmezlik ve Verinin Rasch Modele Uyumu (Anderson Olabilirlik Testi )

İyi bir ölçüm aracı için, madde zorluk ölçümleri, madde zorluklarını kalibre etmek için kullanılan örneklemden değişmez olmalı ve kişi yetenek ölçümleri, yetenek ölçümlerini üretmek için kullanılan madde setinden değişmez olmalıdır (Liu, 2020). Wald testi, medyan ham puandan örnekleme iki alt örneğe ayırır ve Z testi kullanarak iki alt örneklemden elde edilen madde güçlük ölçümlerinin istatistiksel olarak aynı olup olmadığını test eder. Anderson’ın Olabilirlik Testi (Anderson’s Likelihood Ratio Test), Rasch Modeli gibi modellerin veriye uyumunu test etmek için kullanılan istatistiksel bir yöntemdir. Bu test, özellikle veri setinin modele ne kadar uygun olduğunu anlamak amacıyla kullanılır ve modelin geçerliliğini değerlendirmek için önemlidir (Bond ve Fox, 2015). Tablo 3, Wald testi ve Anderson olabilirlik test sonuçlarını göstermektedir.

**Tablo 3. Wald Testi ve Anderson Olasılık Test Sonuçları**

Birinci Analiz			İkinci Analiz			Üçüncü Analiz		
Wald test - medyan			Wald test - medyan			Wald test - medyan		
Soru	z-ist	p	Soru	z-ist	p	Soru	z-ist	p
s1	0,154	0,877	s1	1,249	0,212	s1	0,819	0,413
s2	1,319	0,187	s2	1,385	0,166	s2	1,768	0,077
s3	0,463	0,644	s3	1,385	0,166	s3	1,140	0,254
s4	-1,331	0,183	s4	-2,262	0,024	s5	-0,798	0,425
s5	-0,650	0,516	s5	-0,702	0,483	s8	0,093	0,926
s6	-2,926	0,003	s8	-0,066	0,947	s9	-0,892	0,373
s7	-2,081	0,037	s9	-0,259	0,796	s10	-1,232	0,218
s8	0,871	0,384	s10	-0,892	0,372	s11	0,034	0,973
s9	-0,848	0,396	s11	0,005	0,996	s12	-0,396	0,692
s10	-0,456	0,649	s12	-0,112	0,911	s13	-0,202	0,840
s11	0,020	0,984	s13	0,248	0,804	s14	0,329	0,742
s12	0,804	0,422	s14	1,438	0,151	s15	-0,524	0,600
s13	0,105	0,916	s15	-0,899	0,369	s16	-0,024	0,981
s14	1,129	0,259	s16	-0,016	0,987	s20	0,654	0,513
s15	-1,303	0,193	s20	1,059	0,290			
s16	0,169	0,866						
s19	3,728	0,000						
s20	2,825	0,005						
Andersen's LR test			Andersen's LR test			Andersen's LR test		
Değer (df)		p değeri	Değer (df)		p değeri	Değer (df)		p değeri
Likelihood ratio 42,5(17)		< ,001	15,2(14)		0,368	8,52(13)		0,808
s6, s7 ve s19 çıkartılır			s4 çıkartılır					

Öncelikle tüm maddelere (s17 ve s18 hariç) Wald testi uygulandı, s6, s7 ve s19 sorularının p değeri 0.05'ten küçük olduğundan, değişmezlik varsayımı bu maddeler için ihlal edilmiştir. Bu madde, farklı gruplar için farklı davranıyor olabilir. Ayrıca, Andersen Olabilirlik testine göre de p değeri 0.05'ten küçük olduğu için, Rasch Modeli veriye uymamaktadır. Bu nedenle bu sorular da analiz dışı bırakılıp ikinci bir Wald testi uygulandı. Burada da s4 ün değişmezlik varsayımını ihla ettiği anlaşıldı. Andersen Olabilirlik testine göre de p değeri 0.05'ten büyük olduğu için, Rasch Modelinin veriye uyduğu bulundu. Rasch modelinin veriye uyumunu artırmak ve değişmezlik varsayımını ihlal eden s4 de analiz dışı bırakılmıştır. Üçüncü Wald testi ve Anderson olabilirlik testleri gereğince, Rasch modelin veriye uyduğu (0,808) ve değişmezlik varsayımını ihlal eden soruya rastlanmadığı için Rasch analizine 14 soru ile devam edilecektir.

### 3.3. Test ve Model Güvenirliği

Güvenirlik, çeşitli istatistiksel katsayılar kullanılarak değerlendirildi. Cronbach  $\alpha$  ve McDonald's  $\omega$  verilerin bir ölçüsüdür. Kişi ayırım güvenilirliği, bireylerin ölçekteki performanslarının ne kadar güvenilir bir şekilde ayrıştırılabildiğini gösterir. Kişi Ayırım Güvenirliği, Cronbach  $\alpha$ 'da olduğu gibi, 0,75'in üzerindeki değerler iyi güvenilirliği, 0,9'un üzerindeki değerler ise ölçeğin çok iyi güvenilirliğini gösterir (Wright ve Linacre, 1989; Linacre, 2023).

**Tablo 4. Güvenirlik İstatistikleri**

	<b>Cronbach's <math>\alpha</math></b>	<b>McDonald's <math>\omega</math></b>	<b>Kişi Ayırım Güvenirliği</b>
<b>Ölçüm</b>	0.740	0.743	0.772

Kişi ayırım güvenirligi 0.772, iyi düzeyde kabul edilebilecek bir güvenirlilik düzeyidir. Bu, ölçeğin bireylerin yeteneklerini veya düzeylerini iyi bir şekilde ayırt edebildiğini gösterir. Cronbach's  $\alpha$  (0.740) ve McDonald's  $\omega$  (0.743) değerleri, testin yeterli bir iç tutarlılığa sahip olduğunu göstermektedir. Bu, testin ölçmek istediği yapıyı makul bir şekilde ölçtüğünü ifade eder.

### 3.4. Madde Uyum İstatistikleri

Madde uyum istatistikleri iç uyum (infit MNSQ) ve dış uyum (outfit MNSQ) olmak üzere iki tip istatistiki veri sağlamaktadır. MNSQ, ortalama kare artıkları anlamına gelir. Infit, yetenekleri madde zorluğuna yakın olan öğrencilere daha fazla ağırlık vererek ağırlıklı ortalama kare artıkların ortalaması olarak bulunur. Outfit ise, tüm öğrencilerin eşit şekilde ağırlıklandırılmasıyla ortalama kare artıkların basit ortalaması şeklinde elde edilir. İnfit test maddelerine verilen yanıtlara duyarlı bilgi ağırlıklı uyum istatistiği iken, outfit aykırı değerlere duyarlı uyum istatistiğidir. Nokta seri korelasyonu ise, yanıtlar ile öğrencilerin ölçümleri arasındaki korelasyonu ifade eder. Yüksek korelasyon değeri, bir maddenin yanıtlayanların yeteneklerini ayırt edebildiği anlamına gelir (Linacre, 2023). Dolayısıyla, nokta seri korelasyonuna ilişkin değerlerin 0,30'da büyük olması arzu edilmektedir (Bond ve Fox, 2015). Diğer yandan, Othman vd. (2014) arzu edilen ölçüm değerlerini zayıf ve küçük korelasyon (<0,35), orta ve makul korelasyon (0,36-0,67), güçlü ve yüksek korelasyon (0,68-1,00) halinde kategorilere de ayırmışlardır. Model veri uyumu iyi olan bir sorunun, ikili veri için 0,5-1,7 aralığında infit ve outfit değerlerine sahip olması gerekmektedir (Bond ve Fox, 2015). Tablo uyum istatistiklerini ve nokta seri korelasyonu sunar.

**Tablo 5. İç Uyum, Dış Uyum ve Nokta Seri Korelasyon Değerleri**

<b>Soru</b>	<b>İç Uyum</b>	<b>Dış Uyum</b>	<b>Nokta Seri Korelasyon</b>
<b>s1</b>	0.994	1.139	0.336
<b>s2</b>	1.045	1.077	0.401
<b>s3</b>	1.007	1.043	0.399
<b>s5</b>	0.952	0.823	0.345
<b>s8</b>	0.947	0.965	0.407
<b>s9</b>	0.960	0.972	0.453
<b>s10</b>	0.957	0.927	0.435
<b>s11</b>	0.975	0.951	0.381
<b>s12</b>	1.009	1.126	0.455
<b>s13</b>	1.001	0.955	0.393
<b>s14</b>	1.003	0.995	0.411
<b>s15</b>	0.984	0.918	0.417
<b>s16</b>	1.009	1.048	0.440
<b>s19</b>	1.043	1.123	0.371
<b>s20</b>	1.044	1.009	0.416

Tablo 5 incelendiğinde iç uyum değerlerinin 0,95-1,05 ve dış uyum değerlerinin 0,82- 1,14 aralığında bulunduğu gözlenmektedir. Tüm soruların hem dış uyum hem de iç uyum değerleri kabul edilebilir aralık dâhilindedir. Dolayısıyla elde edilen sonuçlar maddelerin ölçüm için uyum içinde olduğunu gösterir. Ayrıca, Tablo 5'de yer alan nokta seri korelasyon verileri incelendiğinde tüm değerlerinin 0,34-0,45 aralığında değerler aldıkları belirlenmiştir. Test öğrencilerin yeteneğini ayırt edebileceğine işaret etmektedir.

Varsayımlar dikkate alındığında, Rasch model verileri, iç uyum, dış uyum, tek boyutluluk, değişmezlik, verinin Rasch modele uygunluk ve nokta ölçüm korelasyon istatistikleri bağlamında tam bir uyum içinde ve geçerli olduğu görülmüştür.

### 3.5. Madde Parametre Tahminleri

Madde zorluk parametresini elde etmenin amacı, belirli bir yetenek seviyesinde bir soruya doğru cevap verme şansını belirlemektir. Soru zorluğu parametresi logit birimleri ile ifade edilir. İyi soruların zorluk aralığı -2.0 ila +2.0 logit arasındadır (Hambleton ve Swaminathan, 1985). Bir soru maddesi, +2.00 logit'in üzerinde bir zorluk endeksine sahipse çok zor bir soru olarak kabul edilir; bir soru maddesi -2.0 logit'in altında bir zorluk endeksine sahipse çok kolay bir soru olarak kabul edilir. Bir soru maddesi değeri (maddeyi ölçme)  $> 1$  ise çok zor olarak kategorize edilir; zor ise  $0.5 \leq b < 1$ ; orta  $-0.5 \leq b < 0.5$ ; kolay  $-0.5 \leq b < -1$ ; ve çok kolay ise  $b \leq -1$ . (Adedoyin ve Mokobi, 2013:1001). Soruların zorluk parametresi tahminleri Bootstrap Ortak Maksimum Olabilirlik tahmin yöntemi kullanılarak elde edilmiştir. Tablo 6'da ayırt edicilik düzeyi-yetenek ve zorluk parametre tahminleri ve standard hataları verilmiştir.

**Tablo 6. Parametre Tahminleri ve Zorluk Seviyesi**

	Yetenek	Zorluk	Std Hata	Zorluk Derece
s1	0,851	-0,857	0,120	Kolay
s2	0,652	0,504	0,093	Zor
s3	0,726	0,071	0,098	Orta
s5	0,903	-1,436	0,143	Çok Kolay
s8	0,797	-0,409	0,107	Orta
s9	0,669	0,409	0,094	Orta
s10	0,735	0,013	0,099	Orta
s11	0,820	-0,587	0,112	Kolay
s12	0,540	1,109	0,090	Çok zor
s13	0,772	-0,232	0,103	Orta
s14	0,717	0,128	0,097	Orta
s15	0,748	-0,066	0,100	Orta
s16	0,602	0,781	0,091	Zor
s20	0,640	0,572	0,092	Zor

Tablo 6'daki maddeler, yetenek, zorluk, zorluk standart hatası açısından analiz edilmiştir. Ek olarak, her madde için bir zorluk seviyesi belirlenmiştir. Sorular arasından çözülmesi en kolay soru "s5" (-1.436), çözülmesi en zor soru "s12" (1.109), kolay olan sorular "s1" ve "s11", zor sorular "s2", "s16", ve "s20" ve zorluk açısından dengeli sorular olarak görülen sorular "s3", "s8", "s9", "s10", "s13", "s14", ve "s15" sorulardır. Çok kolay sorular daha düşük yetenek seviyesine sahip bireyler için uygundur ve soruların %7,14'ünü, çok zor sorular ise, daha yüksek yetenek seviyesine sahip bireyleri ayırt etmek için kullanışlıdır ve yine soruların %7,14'nü oluşturur. "s5" sorusu çok kolay bir soru ve düşük yetenek seviyesindeki bireyler için uygundur. "s12" sorusu çok zor ve daha yüksek yetenek seviyesine sahip bireyleri değerlendirmek için tasarlanmıştır. Soruların %14,2'si kolay, %50'si orta ve %21,4'ü zor sorulardan oluşmaktadır. Bu maddeler, farklı zorluk seviyelerine göre dengeli bir şekilde dağıtılmıştır. Zorluk seviyeleri geniş bir aralıkta dağıldığı için, bu maddeler testin her seviyedeki bireyleri değerlendirme kapasitesine sahiptir.

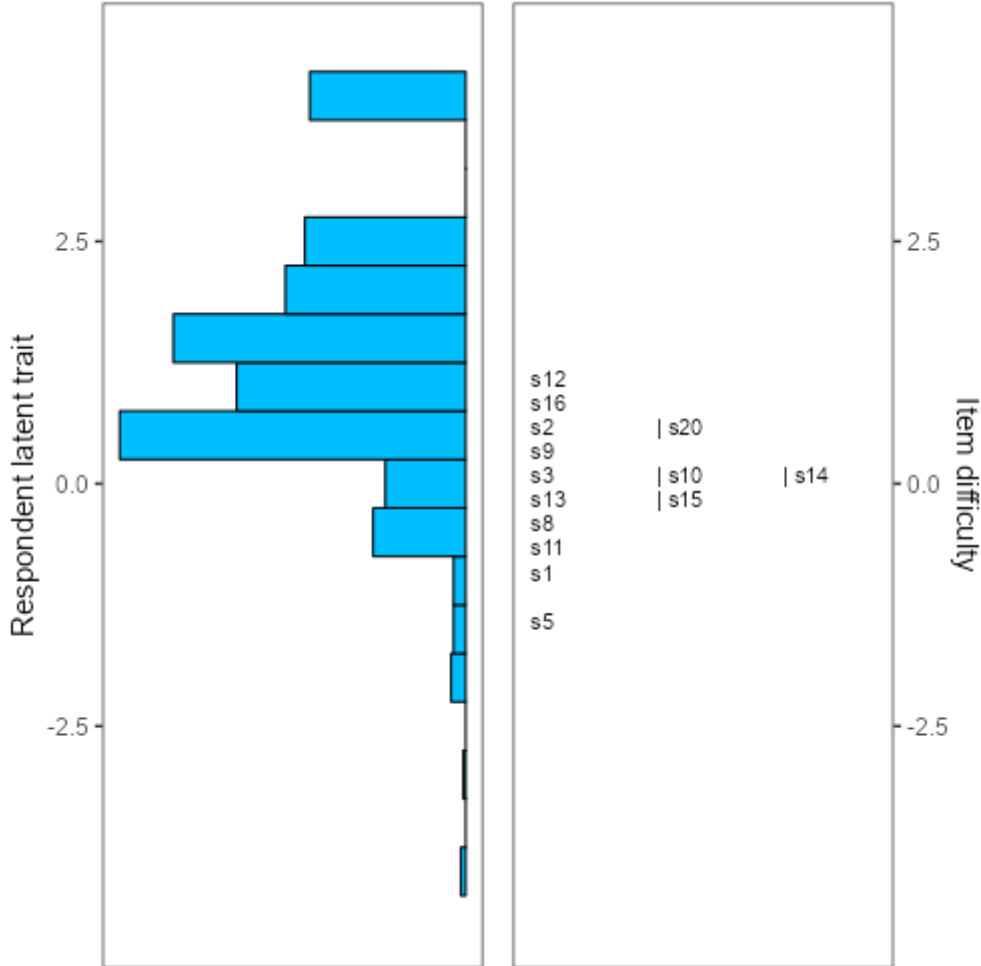
Testin ayırt ediciliği, maddelerin bireylerin yetenek düzeylerini ne kadar iyi ayırt edebildiğini ifade eder. Daha yüksek yetenek değerine sahip maddeler, bireyler arasında yetenek farklılıklarını daha iyi ayırt eder. Yetenek değerleri, genel olarak 0.540 ile 0.903 arasında değişmektedir. Bu aralık, testin sorularının ayırt edicilik açısından makul düzeyde olduğunu göstermektedir. "s5" (0.903), bireylerin yetenek düzeylerini en iyi ayırt edebilen maddedir. "s12" (0.540) diğerlerine göre ayırt edicilikte daha zayıftır. Tablo 6'da genel olarak yetenek değerlerinin yüksek olması, testin bireyler arasında yetenek farklılıklarını iyi ayırt edebileceğini göstermektedir. Zorluk düzeyleri de dengeli dağıldığı için, test hem düşük hem de yüksek yetenek düzeyindeki bireyleri ayırt etme kapasitesine sahiptir. "s12" ve "s16" gibi düşük ayırt ediciliğe sahip maddeler olduğundan yeniden değerlendirilmesi faydalı olabilir.

### 3.6. Wright'ın Haritası

Wright haritası, Rasch analizinin önemli bir çıktısı olup, hem bireylerin yetenek düzeylerini hem de maddelerin güçlük seviyelerini aynı ölçek üzerinde görselleştirmek için kullanılır. Bu görsel, testin ne kadar iyi dengelendiği ve katılımcıların yetenek dağılımıyla maddelerin güçlük seviyelerinin uyumlu olup

olmadığını değerlendirmek için kritik bir araçtır. (Boone, 2016). Şekil 1 Wright haritasını göstermektedir. Grafiğin alt kısmı logit cinsinden Rasch ölçeğidir, sol taraftaki dikey çizgi öğeleri sunar, histogramların üst paneli alt taraftaki Rasch ölçeği boyunca deneklerin yetenek tahminlerinin frekans dağılımını gösterir ve son olarak noktalar karenin içinde alttaki Rasch ölçeğindeki madde zorluklarının yerleri gösterilir.

Şekil 1 Wright Map



Haritanın sol tarafında katılımcıların yetenek düzeyleri yer almaktadır. Görüldüğü üzere, katılımcıların büyük bir kısmı ortalama yetenek seviyesinde toplanmıştır. Bu, testin hedef kitlesinin yetenek açısından homojen veya daha güçlü bir gruba sahip olduğunu gösterir. Negatif tarafa düşen bireylerin sayısı oldukça azdır, bu da düşük yetenek düzeyine sahip katılımcıların teste bulunmadığını gösterir. Haritanın sağ tarafında maddelerin güçlük seviyeleri yer alır. Maddeler genellikle ortalama güçlük seviyesine yakın bir dağılım göstermektedir, ancak bazı maddeler (örneğin, s5) oldukça kolaydır, diğerleri (örneğin, s12, s16) daha zordur. Maddelerin geniş bir zorluk derecesine sahip olması, testin farklı yetenek düzeylerindeki bireyler için uygun bir ölçüm sunduğunu gösterir. Ancak bazı maddelerin zorluk seviyesinin katılımcıların yetenek seviyeleriyle uyum sağlamadığı gözlemlenir. Haritanın genel yapısı, maddelerin güçlük seviyelerinin katılımcıların yetenek seviyeleriyle kısmen uyumlu olduğunu göstermektedir. Ancak, yetenek dağılımının üst ucunda yer alan bazı katılımcılar için yeterince zorlayıcı madde bulunmadığı dikkat çekmektedir. Bu durum, testin yüksek yetenek düzeyindeki bireyler için ayırt edici gücünün sınırlı olduğunu gösterir.

#### 4.SONUÇ VE ÖNERİLER

Rasch analizi, sınav sorularının öğrencilerin yeteneklerine göre uygun zorluk seviyesinde olup olmadığını belirlemeye yardımcı olur ve dengeli bir değerlendirme sağlar. Sınavın genel kalitesini iyileştirmek için değişiklik veya kaldırma gerekebilecek uyumsuz soruları belirleyerek test sorularının güvenilirliğini ve



geçerliliğini değerlendirir. Model, her sorunun çakışma olmadan değerlendirmeye anlamlı bir şekilde katkıda bulunmasını sağlayarak soru uygunluğunu kontrol eder. Rasch analizi, soruların öğrencinin yetenek düzeyine göre uyarlandığı ve öğrenme deneyimini geliştiren uyarlanabilir test ortamları oluşturmak için kullanılır.

Bu çalışmanın amacı, 1. sınıfa başlayacak öğrencilerin İngilizce muafiyet sınavının ölçme geçerliliği ve güvenilirliğini geliştirmek amacıyla Rasch analizi yönteminin uygulanabilirliğini göstermek ve sınav sorularının zorluk seviyeleri ile ayırt edicilik özelliklerini değerlendirmedeki önemini ortaya koymaktır. Çalışmada testin Rasch modele uygunluğu, Rasch model varsayımları titizlikle el alınmıştır. Tek boyutluluk varsayımı için faktör analizi sonucunda tek boyutluluk sağlanmış ancak iki sorunun (s17 ve s18) testi temsil etme gücü düşük çıkmış ve sorular arasından çıkarılmıştır. Martin Løef testi ile de tek boyutluluğun sağlandığı görülmüştür. Değişmezlik varsayımı için Wald testi uygulaması sonucunda s6, s7 ve s19 sorularının değişmezlik varsayımı bu maddeler için ihlal edilmiştir ve sorular arasından çıkarılmıştır. Andersen Olabilirlik testine göre de, bu şekilde Rasch Modeli veriye uymamaktadır. İkinci Wald test ve Andersen Olabilirlik testine göre de s4 sorusu, sorular arasından çıkarılmıştır. Üçüncü Wald testi ve Anderson olabilirlik testleri gereğince, Rasch modelin veriye uyduğu (0,808) ve değişmezlik varsayımını ihlal eden soruya rastlanmadığı için Rasch analizine 14 soru ile devam edilmiştir. Bulgular, tüm soruların 0,34 ile 0,45 arasında değişen bir nokta seri korelasyon katsayısına, iç uyum ve dış uyum değerlerinin 0,82 ile 1,14 arasında, madde güçlük değerlerinin ise -1,44 ile 1,11 arasında değiştiğini gösterdi. Testin Cronbach's  $\alpha$  katsayısı 0.740 ve McDonald's  $\omega$  katsayısı 0.743 olarak bulundu. Kişi ayırım güvenilirliği 0.772 ölçeğin bireylerin yeteneklerini veya düzeylerini iyi bir şekilde ayırt edebildiği ele edildi. Çalışmanın sonuçları, anketin Rasch modeli gerekliliklerine iyi uyum sağladığını gösterdi. Modelin güvenilirliği ve değişmezlik analizleri, testin genel olarak güvenilir bir ölçüm aracı olduğunu desteklemektedir. Ancak, belirli alt gruplar arasında ölçüm değişmezliğinin sağlanıp sağlanmadığı konusu, daha derinlemesine bir analiz gerektirmektedir. Bu durum, testin farklı öğrenci grupları için uygulanmalı ve değişmezlik varsayımının eşit düzeyde geçerli olup olmadığı sınımlıdır.

Rasch modelinin sınav geçerliliği üzerindeki etkisi, literatürde sıklıkla ele alınan bir konudur. Bond ve Fox (2007), Rasch analizinin, sınavın içsel geçerliliğini artırma konusunda önemli bir araç olduğunu belirtmiş ve bu analizle sınav maddelerinin, öğrencilerin beceri seviyelerine uygun bir şekilde ayarlandığını vurgulamıştır. Bu çalışmada da benzer şekilde, Rasch analizi, sınavın her bir maddesinin zorluk düzeylerini ve ayırt ediciliğini belirleyerek geçerliliği artırmıştır. Ancak, Wright ve Masters (1982), modelin bazen sınav sorularının dilini veya içeriğini yeterince dikkate almadığını ifade etmiştir. Bu noktada, çalışmamızda da gözlemlenen bulgular, bazı soruların geçerliliğini artırmak amacıyla revize edilmesi gerektiğini göstermektedir.

Draugalis vd. (2004) Rasch analizinin müfredatın etkinliğini değerlendirme ve öğrenci performansının analizi konusundaki gücünü ortaya koymaktadır. Benzer şekilde, bu çalışma da sınav sorularının geçerliliğini ve güvenilirliğini artırmak için Rasch modelini kullanarak soruların zorluk ve ayırt edicilik düzeylerini belirlemiş ve sınavın kalitesini iyileştirmiştir. Her iki çalışma da Rasch modelinin, sınavın iç tutarlılığını artırma ve öğrenciler arasındaki farklılıkları doğru şekilde yansıtmaya kapasitesini vurgulamaktadır. Azrilah vd. (2008), Rasch modelinin ölçüm araçlarının yapısını doğrulamak için kullanıldığını ve bu modelin bir tutumun uzman düzeyindeki mutabakatını anlamada daha iyi bir keşif derinliği sağladığını belirtmiştir. Çalışmada, Rasch analizi ile yapılan değerlendirmeler, tutumların doğru bir şekilde ölçülmesine olanak sağlamıştır. Bu bağlamda, mevcut çalışmada da Rasch modelinin, sınav sorularının doğru ölçülmesini ve öğrencilerin becerilerinin doğru bir şekilde değerlendirilmelerini sağladığı görülmektedir. Ancak, Oliveira vd. (2012)'nin bulguları, Rasch modelinin yalnızca öğrenci performansını değil, aynı zamanda soru türlerinin (örneğin cloze testleri) doğruluğunu ölçmek için de uygun olduğunu göstermektedir. Bu çalışma, soruların türüne ve ölçülen becerilere bağlı olarak Rasch modelinin esnekliğini ve adaptasyon kabiliyetini ortaya koymaktadır. Zulkifli vd. (2010), Malezya Kebangsaan Üniversitesi'nde doğrusal cebir dersine ilişkin çok amaçlı sorularda ikili Rasch modelini kullanarak madde yanıtının 0-1 puanlamasını analiz etmiştir. Benzer şekilde, bu çalışmada da, Rasch modelinin sınav sorularını ayırt edici bir şekilde değerlendirme yeteneği kullanılmıştır. Zulkifli vd. (2010)'nin çalışması, 0-1 puanlamasını kullanarak soruların daha net bir şekilde analiz edilmesine olanak tanımaktadır. Bu durum, çalışmamızda da benzer şekilde, her bir sınav sorusunun zorluk ve ayırt ediciliğinin belirlenmesi için ikili modelin faydalı olduğuna işaret etmektedir. Zulkifli vd. (2012), Rasch modelinin hem öğrenci yeteneğini hem de soru geçerliliğini ölçmek için kullanıldığını ve bu modelin öğrenci başarısını uygun soru setleriyle artırma potansiyeline sahip olduğunu belirtmiştir. Bu çalışma, Rasch modelinin, belirli derslerdeki hedeflere uygun soru setlerinin oluşturulmasına yardımcı olduğunu vurgulamaktadır. Bu bulgu, mevcut çalışmanın bulgularıyla

paralellik göstermektedir; çalışmamızda da soruların zorluk seviyelerinin optimize edilmesi ve soruların öğrencilerin başarı seviyelerini daha iyi yansıtabilecek şekilde geliştirilmesi gerektiği sonucuna varılmıştır. Boone (2016), Rasch analizinin test geliştirmeye rehberlik edebileceğini, puan anlamlarını açıklayabileceğini ve farklı test formlarının oluşturulmasına olanak sağlayabileceğini ifade etmiştir. Bu çalışma, Rasch modelinin sınavların tasarımına nasıl rehberlik edebileceğini ve sınav formlarının birbirleriyle karşılaştırılabilir olmasını sağladığını göstermektedir. Benzer şekilde, bu çalışmada da Rasch modeli kullanılarak sınav sorularının geçerliliği ve güvenilirliği artırılmış, sınavın her düzeydeki öğrenci için uygunluğu sağlanmıştır.

Chen ve Zhou (2018), Çin'deki Ulusal Matrikülasyon İngilizce Testi (NMET) üzerine yaptıkları çalışmada, Rasch modelini cloze testi analizinde kullanmışlardır. Bu çalışmanın bulguları, Rasch modelinin dil becerilerini ölçen testlerde de güçlü bir araç olduğuna işaret etmektedir. Benzer şekilde, Gan vd. (2018), Çin'deki TEM4 testini incelemiş ve Rasch modelinin testin geçerliliğini artırmadaki etkinliğini göstermiştir. Bu iki çalışma, Rasch modelinin dil becerileri ve sınavların genel geçerliliğini değerlendirmede önemli bir araç olduğuna dikkat çekmektedir. Bu, çalışmamızda da gözlemlenen bulgulardan biridir; Rasch analizi, sınavın her bir maddesinin ne kadar etkili ve geçerli olduğunu ortaya koymuştur.

Tunç (2023), Türkiye'deki Rasch temelli ölçme aracı geliştirme tezlerinin, modelin temel varsayımlarını ve analizlerini raporlarken tutarsızlıklar gösterdiğini ve bu tekniklerin uygulanmasında daha sistematik bir yaklaşım gerekliliğini vurgulamıştır. Bu eleştiri, mevcut çalışmamızın bulgularıyla da uyumludur. Rasch modelinin uygulaması, özellikle düşük örneklem gruplarında veya küçük sınav setlerinde dikkatli bir şekilde yapılmalıdır. Ayrıca, bu çalışma da modelin daha sistematik bir şekilde uygulanması gerektiğine işaret etmektedir. Rasch analizinin tüm varsayımlarının doğru bir şekilde raporlanması ve modelin her aşamasının titizlikle gözden geçirilmesi gerektiği açıktır.

Soruların bazıları (örneğin, s5) kolay bulunurken, bazıları (örneğin, s12 ve s16) oldukça zordur. Testin daha dengeli hale gelmesi için, katılımcıların yetenek seviyeleriyle daha iyi eşleşen maddeler eklenebilir. Wright Harita, test maddelerinin çoğunlukla katılımcıların yetenek seviyelerine uygun olduğunu, ancak aşırı yüksek veya düşük yetenek düzeyine sahip bireyler için uygun maddelerin eksik olduğunu göstermektedir. Bu durum, testin genel geçerlik ve güvenilirlik açısından iyileştirilmesi gerektiğini gösterir. Maddeler genelde orta güçlük seviyesinde toplanmış olsa da, testin daha geniş bir yetenek aralığını kapsamaya için hem daha kolay hem de daha zor maddeler eklenmelidir. Test, ortalama yetenek düzeyine sahip bireyler için oldukça uygun görünmektedir. Ancak, düşük ve yüksek yetenek düzeyine sahip bireyler için uygun maddelerin artırılması testin gücünü artıracaktır. Üst yetenek düzeyinde yer alan katılımcılar için yeterince zorlayıcı maddelerin eksikliği dikkat çekicidir. Daha karmaşık ve ayırt edici maddeler eklenmesi önerilebilir. Testin tüm yetenek düzeylerinde güvenilir ölçüm sunabilmesi için soru setinin daha dengeli bir şekilde tasarlanması ve analizlerin yeniden yapılması faydalı olacaktır. Rasch modeline ek olarak, diğer Madde Tepki Teorisi (MTT) modelleri kullanılarak testin performansı farklı açılardan değerlendirilebilir. Bu, testin güçlü ve zayıf yönlerini daha kapsamlı bir şekilde anlamaya yardımcı olur ve ileriki çalışmalar için önerilebilir.

Bu analizle, testin hem içeriği hem de uygulama sonuçları hakkında önemli iyileştirme alanları belirlenmiştir. Özetle, Rasch Modeli, öğretmenlere ve araştırmacılara, klasik test teorisine kıyasla öğrencinin yeteneği ve test soruları arasındaki ilişki hakkında daha zengin yorumlar sağlamaları için bir yol sağlar. Rasch analizi, sınavlarının kalitesini artırmayı amaçlayan eğitimciler için onu olmazsa olmaz bir araç haline getirir.

## KAYNAKÇA

Adedoyin, O. ve Mokobi, T. (2013). Using IRT Psychometric Analysis in Examining The Quality of Junior Certificate Mathematics Multiple Choice Examination Test Items. *International Journal of Asian Social Science*, 3(4), 992–1011, <https://doi.org/10.1177/2158244019832706>.

Azrilah, A.A., Azlinah, M., Azami, Z., Sohaimi, Z., Hamzah, A.G. ve Saidfudin, M. (2008). Evaluation of Information Professional Competency Face Validity Test Using Rasch Model. *5th WSEAS/IASME International Conference on Engineering Education (EE'08)*, Heraklion, Greece, 396-403.

Baghaei, P. ve Ravand, H. (2019). Method Bias in Cloze Tests as Reading Comprehension Measures. *Sage Open*, 9(1), <https://doi.org/10.1177/2158244019832706>.

Bond, T. G., ve Fox, C. M. (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Routledge, New York.

- Boone, W.J. (2016). Rasch Analysis for Instrument Development: Why, When, and How? *CBE Life Sciences Education*, 15 (4), .
- Chen, Y. and Zhou, R. (2018). The application of Rasch Model in the quality analysis of cloze items in NMET. *Foreign Language Testing and Teaching* 1, 39-47, 64
- Dhyaaldian, S. M. A., Al-Zubaidi, S. H., Mutlak, D. A., Neamah, N. R., Albeer, A. A. M. A., Hamad, D. A., Al Hasani, S. F., Jaber, M. M. ve Maabreh, H. G. (2022). Psychometric Evaluation of Cloze Tests with the Rasch Model. *International Journal of Language Testing*, 12(2), 95-106.
- Draugalis, J.R. ve Jackson, T.R. (2004). Objective Curricular Evaluation: Applying the Rasch Model to a Cumulative Examination. *American Journal of Pharmaceutical Education* 68 (2): 1-12. <http://archive.ajpe.org/aj6802/aj680235/aj680235.pdf>. [12 November 2011].
- Embretson, S. E., ve Reise, S. P. (2000). *Item Response Theory for Psychologists*. Lawrence Erlbaum Associates Publishers.
- Gan Q., Yang, Y. ve Li, Y. Q. (2018). Rasch Model Based Comparative Study on Validity between Multiple-Choice Cloze Questions and Banked Cloze Questions in TEM-4. *Journal of Chengde Petroleum College*, 20(4), 49-53
- Hambleton, R. K., Swaminathan, H., ve Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Sage Publications, Inc.
- Linacre, J. M. (2023). A User's Guide to Winsteps? Ministep: Rasch-Model Computer Programs. Program Manual 5.5.1. <https://www.winsteps.com/winman/copyright.htm>
- Mair, S., Jones, A., Ward, J., Christie, I., Druckman, A., ve Lyon, F. (2018). A Critical Review of The Role of Indicators in Implementing the Sustainable Development Goals. *Handbook of Sustainability Science and Research*, 41-56.
- Mokshein, S. E., Ishak, H. ve Ahmad, H. (2019). The Use of Rasch Measurement Model iEnglish Testing. *Cakrawala Pendidikan: Jurnal Ilmiah Pendidikan*, 1(1), 16-32.
- Oliveira, K., D. S., Boruchovitch, E. ve Rueda, F (2012). Reading Comprehension: Differential Item Functioning Analysis of a Cloze Test. *Paper Research*, 25(2), 221-229.
- Rasch G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Danish Institute for Educational Research, Copenhagen.
- Richichi, R. (1996). An Item Response Theory Analysis of Multiple-Choice Items Chosen at Random From a Publisher's Test Bank. *Presentation at the 27th annual Northeastern Educational Research Association Conference*, October, Ellenville, 2205-2214.
- Tunç, E. B. (2023). A Review of Measurement Tools Developed and Adapted Based on the Rasch Model. *İnsan Ve Sosyal Bilimler Dergisi*, 6(2), 249-275. <https://doi.org/10.53048/johass.1369336>.
- Wright, B. D. ve Masters, G. N. (1982). *Rating Scale Analysis*. MESA Press.
- Wright, B. D. ve Linacre, J. M. (1989). Observations are Always Ordinal; Measurements, However, must Be Interval. *Archives of Physical Medicine and Rehabilitation*, 70(12), 857-860.
- Zulkifli, M.N., Mohd H. O., Noorhelyna, R., Fadiyah, M.A. ve Izamarlina, A. (2010). How Good was the Test Set Up? From Rasch Analysis Perspective. *Regional Conference on Engineering Education ve Research in Higher Education (RCEE ve RHED)*, 7-9 June, Kuching.
- Zulkifli, M. N., Mohd H. J., Nur A. I., Haliza O., Izamarlina A. ve Mohd H. O. (2012) Reliability Analysis of Examination Questions in a Mathematics Course Using Rasch Measurement Model. *Sains Malaysiana*, 41 (9). 1171-1176.