



Bolu Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi (BAİBÜEFD)

Bolu Abant İzzet Baysal University
Journal of Faculty of Education

2023, 23(4), 2160–2186. <https://dx.doi.org/10.17240/aibuefd.2023.-1189084>



Sembolik ve Sözel Biçimi Kullanma Öz Yeterlik Ölçeğinin Psikometrik Özelliklerinin Rasch Modeli ile İncelenmesi

Investigation of Psychometric Properties of Using Symbolic and Verbal Forms Self-Efficacy Scale with Rasch Model

Erdem ÇEKMEZ¹ ID, Rümeysa CEVAHİR BOLAT² ID

Geliş Tarihi (Received): 19.10.2022

Kabul Tarihi (Accepted): 05.12.2023

Yayın Tarihi (Published): 15.12.2023

Öz: Bu çalışmada, matematik öğretmeni adaylarının matematiksel ifadelerin farklı biçimleri arasında tercüme yapma hususunda öz yeterliklerini ölçmek için Klasik Test Teorisi çerçevesinde geliştirilmiş bir ölçeğin psikometrik özelliklerinin Rasch Teorisi perspektifinden incelenmesi amaçlanmıştır. Bu amaçla ölçek 252 matematik öğretmeni adayının oluşturduğu bir örneklem üzerinde uygulanarak elde edilen veriler ile maddelerin uyum istatistikleri, kişiler ile maddelere ilişkin güvenilirlik ve ayırt edicilik indeksleri, maddelere yanıt verme hususunda benimsenen kategori yapısının etkinliği, test maddelerinin cinsiyet, coğrafi bölge ve sınıf düzeyi değişkenlerinin oluşturduğu örneklemin alt kümelerinde yanlış gösterme durumu, testin örnekleme hedefleme yeterliği ve testin boyut yapısı incelenmiştir. Elde edilen bulgular örneklemin farklı alt kümelerinde ölçek maddelerinin yanlış göstermediğini, uyum istatistiklerinin kabul edilebilir sınırlar içerisinde olduğunu, ölçeğin Rasch boyutundan başka boyut içermediğini, test maddelerinin öz yeterlik değişkeninin geniş bir aralığını ve kişileri ilgili niteliğe sahiplik açısından kabul edilebilir hassaslıkta ölçebildiğini, testin örnekleme hedeflemede yeterli olduğunu göstermiştir. Ayrıca ölçekte benimsenen kategori yapısının uygun olmadığı ve olumsuz ve olumlu maddelerin zıtlık oluşturduğu ortaya çıkmıştır. Elde edilen bulgulardan hareketle test maddelerine ve kategori yapısına yönelik önerilerde bulunulmuştur.

Anahtar Kelimeler: Matematik dili, matematiksel ifadelerin tercümesi, öz yeterlik, Rasch modeli

&

Abstract: The study aims to examine the psychometric properties of a scale developed within the framework of Classical Test Theory to measure pre-service mathematics teachers' self-efficacy in translating between different forms of mathematical expressions from the perspective of Rasch Theory. For this purpose, the scale was administered to a sample of 252 pre-service mathematics teachers and the data obtained were used to examine the fit statistics of the items, the reliability and discrimination indices of the items with individuals, the effectiveness of the category structure adopted in responding to the items, the bias of the test items in the subsets of the sample formed by gender, geographical region and grade level variables, the adequacy of the test to target the sample and the dimension structure of the test. The findings showed that the scale items did not show bias in different subsets of the sample, the fit statistics were within acceptable limits, the scale did not contain any dimensions other than the Rasch dimension, the test items could measure a wide range of self-efficacy variables and individuals with acceptable precision in terms of having the relevant attribute, and the test was adequate in targeting the sample. In addition, it was revealed that the category structure adopted in the scale was not appropriate and that negative and positive items were contrasted. Based on the findings, suggestions were made for the test items and category structure.

Keywords: Mathematical language, translation of mathematical statements, self efficacy, Rasch model

Atf/Cite as: Çekmez, E. Ve Cevahir-Bolat, R (2023). Sembolik ve sözel biçimi kullanma öz yeterlik ölçeğinin psikometrik özelliklerinin Rasch modeli ile incelenmesi. *Bolu Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 23(4), 2160-2186 doi.org/10.17240/aibuefd.2023.-1189084

İntihal-Plagiarizm/Etik-Ethic: Bu makale, en az iki hakem tarafından incelenmiş ve intihal içermediği, araştırma ve yayın etiğine uyulduğu teyit edilmiştir. / This article has been reviewed by at least two referees and it has been confirmed that it is plagiarism-free and complies with research and publication ethics. <https://dergipark.org.tr/pub/ijaws>

Copyright © Published by Bolu Abant İzzet Baysal University– Bolu

¹ Doç. Dr. Erdem ÇEKMEZ, Trabzon Üniversitesi, Matematik ve Fen Bilimleri Eğitimi Bölümü, erdemcekmez@gmail.com, ORCID: 0000-0001-8684-2820

² Sorumlu Yazar: Arş. Gör. Rümeysa CEVAHİR BOLAT, İstanbul Üniversitesi-Cerrahpaşa, Matematik ve Fen Bilimleri Eğitimi Bölümü, rumeysa.cevahir@iuc.edu.tr, ORCID: 0000-0002-5274-6715

1. GİRİŞ

Matematik dersleri ve ders kitaplarının yanı sıra matematiği barındıran her yerde kullanılan yazılı dil matematik dilidir. Bu dilin kendine has özellikleri (sözcüksel, sembolik ve simgesel) ve yapısı uluslararası bir dil olarak kabul edilmesini sağlar (Uğurel ve Morali, 2010; Yıldırım, 2018). Dünya ölçeğinde farklı ana dile sahip bireylerin aynı sembolleri ve söz dizimi kurallarını kullanarak aynı anlama gelen matematiksel ifadeleri yazabiliyor olmaları matematiğe evrensel bir dil olma özelliğini vermektedir (Uygur-Kabael, 2017). Matematiksel bilginin iletişimi amacıyla oluşturulan matematiksel ifadeler, hem semboller ve notasyonların kullanımıyla sembolik biçimde hem de ana dildeki kelimeler kullanımıyla sözel biçimde ifade edilebilmektedir. Matematik öğrenme ortamlarında öğrenciler ve öğretmenler, matematik dilinin bu iki temsil formları arasında sürekli geçiş yapmaktadır.

Ulusal ve uluslararası resmî kurumlarca hazırlanmış öğretim programlarında öğrencilerin etkili biçimde matematiği bir iletişim aracı olarak kullanabilmesi gerektiği vurgulansa da farklı öğrenim kademelerinde gerçekleştirilen çalışmalarda (Çekmez, 2020; Otterburn ve Nicholson, 1976; Uygur-Kabael, 2017) öğrencilerin matematiksel iletişim becerisinin yeterli düzeyde olmadığı ve öğrencilerin matematik dilini doğru kullanmada sıkıntılar yaşadıkları rapor edilmiştir. Matematik diline ilişkin ulusal alanyazın incelendiğinde öğrencilerin matematik dilini kullanabilme yeterliklerinin nitel yöntemler kullanan (Fırat ve Dinçer, 2018; Uygur-Kabael, 2017), öğrencilerin matematiksel iletişimine, matematik dil kullanımına yönelik öğretmen farkındalıklarına odaklanan (Açıl ve Zeybek, 2017; Kabael ve Baran, 2016), matematik öğretmeni adaylarının matematik dili öz yeterliklerinin öz değerlendirme becerisine etkisini konu edinen (Kabael ve Yayan, 2017) çalışmalar dikkat çekmektedir.

Öğrencilerin matematik dilini kullanma becerilerinin gelişmesinde matematik dilinin kullanıldığı öğretim ortamlarında bulunmaları etkilidir. Öğrencilerin matematik diline maruz kalabilmesi için de matematik öğretmenlerinin derslerinde matematik dilini doğru kullanmaları ve açık veya örtük biçimde matematik dilini vurgulamaları gerekmektedir. Öğretmenlerin matematik dilini vurgulayabilmesinde hem gerekli matematik dili bilgisine sahip olması hem de matematik dilini doğru kullanacağına ve öğrencilerde bu dil gelişimini sağlayabileceğine yönelik öz yeterliği öğretmeni olumlu yönde etkilemektedir. Öğretmenin bir beceriyi öğrencilerine kazandırabileceğine yönelik kendine olan inancı yüksekse, ilgili beceriyi öğretmeye yönelik yüksek motivasyona sahip olur. Bunun sonucu olarak, matematik öğretmenlerinin matematik dilini öğretimde doğru kullanmasını sağlayacak önemli unsurlardan biri matematik dilini kullanmaya yönelik öz yeterliğidir. Bandura öz yeterlik olgusunu, “bireyin belirli kazanımları üretmek için gerekli olan eylemleri organize etme ve yürütme yeteneklerine olan inancı” olarak tanımlanmaktadır (Bandura, 1977, s. 3). Yüksek öz yeterliğe sahip öğretmenler diğer meslektaşlarına kıyasla öğretimde daha istekli ve zorluklara karşı daha dirençlidir (Demirci, 2021; Tschannen-Moran ve Hoy, 2007; Tuchman ve Isaacs, 2011). Öğretmenlerin güçlü öz yeterliğinin olması için matematik öğretmeni adaylarının, lisans eğitim sürecinde matematiğin evrensel bir dil olduğu ile ilgili farkındalığı kazanması, matematik dilinin özellikleri ve kullanımı konusunda bilgilendirilmesi hedeflenmelidir. Böylece bu alanda öz yeterliklerinin gelişmesine katkı sağlayabilecek ortamlar oluşturulabilir. Öğretmen adaylarının bu kazanımlara erişip erişemediğini veya ne düzeyde eriştiğini belirleyebilmede öz yeterlik ölçeği kullanılabilir ve sınıf düzeylerine göre araştırmalar yürütülerek öz yeterliklerinin gelişimi incelenebilir. Yurt içi literatür incelendiğinde matematik öğretmeni adaylarının matematik diline yönelik öz yeterliklerinin ölçülebileceği bir teste rastlanmamıştır. Bu boşluğa katkı sunma amacıyla ikinci sıradaki yazar tarafından birinci sıradaki yazarın danışmalığında tamamladığı yüksek lisans tezi çalışmasında Klasik Test Teorisi yaklaşımıyla bir ölçek geliştirilmiştir (Cevahir, 2022). Son yıllarda eğitim, psikoloji ve sağlık gibi birçok alanda bireylerin sahip oldukları nitelikleri ölçmede Rasch Ölçme Teorisinin modelleri popülerlik kazanmıştır. Eğitim alanından örnek verilecek olursa PISA sınavının analizinde Rasch modelleri kullanılmaktadır (Boone, Staver ve Yale, 2014). Bu değişimin temel sebeplerinden birisi, verilerin modele uyması durumunda yapılan ölçüm sonunda elde edilen ölçülerin eşit aralıklı olmasıdır. Bir diğer sebebi ise, gerçekleştirilen bir Rasch analizi

sonucunda geçerlik ve güvenilirlik hakkında Klasik Test Teorisinin sunabildiğinden daha fazla bilginin edinilmesidir. Buradan hareketle, bu çalışmada Rasch Ölçme Teorisi'nin sunduğu avantajlardan yararlanarak ilgili testin geçerlik ve güvenilirliği hakkında daha detaylı fikir edinmek amaçlanmıştır. Bu amaç doğrultusunda araştırmada ele alınan ana problem, hazırlanmış ölçeğin aşağıdaki sorular çerçevesinde ne düzeyde geçerli ve güvenilir ölçüm gerçekleştirdiğidir. Bu ana problem bağlamında aşağıdaki sorulara cevap aranmıştır.

- 1- Ölçek maddelerinin cevaplandırılması için benimsenen kategori yapısı uygun mudur?
- 2- Kişi ve madde güvenilirlik ile ayırt edicilik indeksleri uygun mudur?
- 3- Maddelerin uyum istatistikleri kabul edilebilir aralıkta mıdır?
- 4- Ölçek örnekleme iyi hedeflemekte midir?
- 5- Maddelerin hiyerarşisi ve dağılımı uygun mudur?
- 6- Ölçek maddeleri yanlılık göstermekte midir?
- 7- Ölçek Rasch boyutundan farklı bir boyut içermekte midir?
- 8- Ölçek yerel bağımsızlık kriterini sağlamakta mıdır?

1.1 Rasch Derecelendirme Ölçeği Modeli

Derecelendirme ölçekleri eğitim araştırmalarında yaygın olarak kullanılmaktadır. Lakin, Klasik Test Teorisi bağlamında bu ölçeklerden elde edilen ham puanların araştırma sorularını cevaplamada kullanılması birçok sınırlılığı beraberinde getirmektedir (Boone vd., 2014). Örneğin, bu ölçeklerden elde edilen verilerde kimi zaman maddelerin puanları normallik şartlarını sağlamayabilmektedir. Böyle bir durumda madde puanlarının ortalamasını hesaplamak uygunsuz, veriler kullanılarak yapılan yorumlar ise yanıltıcı olmaktadır (Neumann vd., 2011). Bunun yanı sıra derecelendirme ölçeklerinden elde edilen puanların eşit aralıklı olduğu çıkarımının yapılması hatalı görülmektedir (Oon ve Subramaniam, 2011).

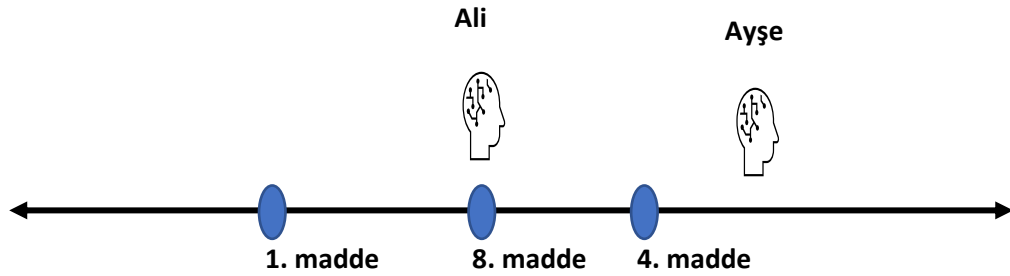
Bu tarz ölçeklerde maddelere yanıt vermek için onaylanabilecek 2'den fazla kategori bulunmakta ve kategoriler genelde 1'den başlayan ardışık tam sayılar ile puanlanmaktadır. Bu puanlama yöntemi ise ardışık kategorilerin ilgili örtük değişken bağlamında ima ettiği değişimin eşit olduğu anlamını taşır. Araştırmacıların bu anlamı sorgusuz olarak kabul etmesi Rasch modelini benimseyen uzmanlar tarafından hata olarak değerlendirilmektedir (Boone ve Staver, 2020). Dolayısıyla bu modeli benimseyen uzmanlar, derecelendirme ölçeklerinden elde edilen verilerin eşit aralıklı (doğrusal) yerine sıralama tipinden muamele görmesi gerektiğini savunurlar. Ayrıca kategorilerde olduğu gibi maddeler arasında da eşit aralığın olmadığını savunmaktadırlar (Boone vd., 2014). Örneğin, bir derecelendirme ölçeğinde ham puanı sırasıyla 60 ve 70 olan iki kişi arasındaki farkın 70 ve 80 olan başka iki kişi arasındaki farka eşit olmadığı görüşündedirler.

Ölçek geliştirmede ve bu ölçeklerden elde edilen performansların nicelleştirilmesinde Rasch modeli kayda değer avantajlar sunmaktadır (Bond ve Fox, 2007). Rasch modeli derecelendirme ölçeklerinden elde edilen puanları sıralama türünden kabul ederek bu puanlardan eşit aralıklı puanlar oluşturur (Boone vd., 2014). Rasch modeli bağlamında derecelendirme ölçeklerinden elde edilen çok kategorili verilerin analizinde Derecelendirme Ölçeği Modeli (Rating Scale Model) kullanılmaktadır (Andrich, 1978). Bu modelde testin tek bir niteliği ölçtüğü kabulü mevcuttur. Diğer bir ifadeyle testte yer alan maddelerin tamamının tek bir örtük değişkenin farklı göstergeleri olduğu kabulü yapılır (Boone ve Staver, 2020). Bu örtük değişken ise testin Rasch boyutu terimi ile isimlendirilir.

Derecelendirme ölçeği modelinin temelini oluşturan matematiksel denklem aşağıdaki gibidir (Linacre, 2002). Denklem içerisinde yer alan F_j ardışık kategorilerin seçilme olasılığının eşit olduğu eşik değerleri, B_n n'inci kişinin ölçüsünü, D_i i'inci maddenin ölçüsünü, P_{nij} ve P_{nij-1} ise kişinin ardışık kategorileri seçme olasılıklarını temsil etmektedir.

$$\log\left(\frac{P_{nij}}{P_{ni(j-1)}}\right) = B_n - D_i - F_j$$

Yukarıda ifade edilen denklemlerle hesaplanan kişilerin ve maddelerin ölçüleri logit adı verilen ölçü birimi cinsinden nicelleştirilmektedir. Rasch modelini açıklamak için yazılan akademik yayınlarda ve bu model temelinde yapılan araştırma raporlarında, araştırmacılar bir örtük değişkenin miktarını temsil eden değişkeni görselleştirmek için sayı doğrusu modelini kullanmaktadır. Oluşturulan bu sayı doğrusunun oryantasyonu matematikteki sayı doğrusu ile paralellik arz eder. Yani doğru üzerinde sağa doğru gittikçe değerler büyümektedir. Hem kişilerin hem de maddelerin ölçüsü aynı birimle (logit) sayısallaştırıldığından, kişiler ve maddeler aynı sayı doğrusu üzerinde konumlandırılabilir. Bu husus Şekil 1’de hayali bir senaryo ile görselleştirilmiştir. Şekil 1’de resmedilen hayali bağlamda, Ayşe’nin ölçülen niteliğe sahiplik miktarı hem Ali’den hem de resmedilen 3 maddenin karşılık geldiği nitelik düzeyinden daha fazladır. Bir anlamda tüm maddeler Ayşe için kolaydır. Ali için ise 4. madde zor, 1. madde kolay ve 8. maddenin ölçüsü tam olarak Ali’nin yetenek ölçüsüne denk gelmektedir.



Şekil 1. Rasch modelinde kişiler ve ölçüler aynı sayı doğrusu üzerinde temsil edilebilmektedir

Bu çalışmada da yer verildiği gibi, bir Rasch analizi sonucunda kişilerin ve maddelerin ölçülerini karşılıklı olarak görselleştirmek, üzerinde yorum yapmak ve tartışmak için Şekil 1’de yer alan doğru dikey olarak konumlandırılır ve kişi-madde haritası (Wright map) olarak isimlendirilen bir çerçeve oluşturulur. Bu çerçevede yer alan dikey doğru dik koordinat sistemindeki *y-ekseni* olarak düşünülebilir. Bu çerçevede kişiler ve maddeler hem kendi içerisinde hem de birbirleriyle karşılaştırılabilir.

2. YÖNTEM

2.1. Araştırmanın modeli

Betimsel olarak nitelendirilen araştırmalar mevcut bir durumu incelemeyi ve durum hakkında değerlendirme yapmayı amaçlamaktadır (Çepni, 2007). Bu araştırma türü içerisinde yer alan tarama yöntemi ise bir niteliğe yönelik kişilerin sahiplik durumunu belirlemeyi amaçlamaktadır (Büyüköztürk vd., 2019). Bu çalışmada, tarama yöntemi ile matematik öğretmen adaylarının matematik diline ilişkin öz yeterliklerini belirlemek için gerçekleştirilen ölçüm geçerlik ve güvenilirlik açısından incelenmiştir.

2.2. Araştırmanın çalışma grubu

Araştırmanın katılımcılarını Marmara ve Karadeniz bölgesinden yer alan 2 eğitim fakültesinin ilköğretim matematik öğretmenliği programında öğrenim görmekte olan toplam 252 öğretmen adayını oluşturmaktadır. Öğrencilerin demografik değişkenler açısından örneklem içerisindeki dağılımı Tablo 1’de sunulmuştur.

Tablo 1*Katılımcıların Demografik Değişkenlere Göre Dağılımı*

Değişken		f	%	Toplam
Cinsiyet	Kadın	210	83	252
	Erkek	42	17	
Coğrafi bölge	Karadeniz	170	67	252
	Marmara	82	33	
Sınıf seviyesi	1. sınıf	97	38	252
	2. sınıf	77	31	
	3. sınıf	44	17	
	4. sınıf	34	14	

2.3. Veri toplama araçları ve süreci

Çalışmada ikinci yazarın birinci yazar danışmanlığında tamamladığı yüksek lisans tezinde oluşturduğu SSBKÖYÖ veri toplama aracı olarak kullanılmıştır (Cevahir, 2022). Ölçeğin geliştirilmesinde izlenen adımlar tez içerisinde detaylıca açıklandığından ve tez erişime açık olduğundan burada yer verilmemiştir. Çalışmanın sonucunda Ek-1’de sunulan 3 faktörlü “Sembolik ve Sözel Biçimi Kullanma Öz Yeterlik Ölçeği (SSBKÖYÖ)” geliştirilmiştir. Ölçeğin ölçmeyi hedeflediği örtük değişken matematiksel ifadelerin sembolik ve sözel formları arasında tercüme yapabilme becerisine ilişkin öz yeterlidir. Ölçekte 8’i olumsuz olmak üzere toplamda 18 madde bulunmaktadır. Ölçekte yer alan maddelere yanıt verilebilecek 5 kategori bulunup Ek-1’de görüleceği üzere kategoriler 1 ile 5 arasında puanlanmıştır. Puanlama gerçekleştirilirken olumsuz maddeler ters sırada puanlanmıştır. Maddeler, elde edilen ham puan arttıkça öz yeterlik niteliğinin arttığı kurgulanarak puanlanmıştır. Veriler çevrimiçi ortamda oluşturulan form ile ilköğretim matematik öğretmen adaylarına iletilmiş ve toplanmıştır.

2.4. Verilerin analizi

İlköğretim matematik öğretmeni adaylarından toplanan veriler Rasch analiziyle incelenerek SSBKÖYÖ’nün psikometrik özellikleri ortaya konulmuştur. Araştırma sorularına cevap vermek için yürütülen Rasch analizi Winsteps® (sürüm: 3.91) yazılımı ile gerçekleştirilmiştir. Rasch analizinden önce toplanan verilerin kalitesini artırmak için yapılan eylemlere alt başlık altında yer verilmiştir.

2.4.1. Veri Kalitesini Arttırmaya Yönelik Analiz Öncesi Gerçekleştirilen İşlemler

Ölçeğin örnekleme uygulanması sonucunda elde edilen ham puanlara ait betimsel istatistikler Tablo 2’de sunulmuştur.

Tablo 2*Ölçekten elde edilen ham puanların betimsel istatistikleri*

N	\bar{x} (%95 güven aralığı)	ss	çarpıklık	basıklık
252	69.14 (68.07, 70.20)	8.61	0.18	-0.17

Rasch modelinde bir testten alınabilecek en yüksek puanı veya en düşük puanı alan kişiler testin yapı geçerliliğini analiz etmede dikkate alınmaması önerilmektedir. Bunun sebebi, bu kişiler için ölçüm hatasının sonsuz olması ve dolayısıyla testin ne kesinlikte işlev gösterdiğini anlamak için faydalı bilgi sunmamasıdır (Boone vd., 2014). Daha açık olarak, eğer bir kişi bir testten maksimum puanı elde etmişse, test o kişinin ilgili örtük değişkene sahiplik düzeyinin ne olduğunu belirleyemez. Çünkü, kişi testteki tüm maddelerin bir bütün olarak işaret ettiği nitelik düzeyinden daha fazlasına sahiptir. Fakat bu fazlalığın miktarı hakkında yorum yapmak mümkün değildir. Aynı durum minimum puan elde eden kişi için de geçerlidir. Bu gerekçeden ötürü verilerin analizi gerçekleştirilmeden önce testten maksimum ve minimum puan alan bireylerin mevcut olup olmadığı incelenmiştir. Yapılan incelemede minimum puan alan kişinin olmadığı, maksimum puan (90) alan 3 kişinin mevcut olduğu belirlenmiştir. Bu 3 kişi veri setinden çıkartılmıştır.

Bir ölçeğin uygulanması sonucunda elde edilen verilerden doğru yorumlar ve çıkarımlar yapılabilmesinin kriterlerinden biri katılımcıların ölçek maddelerini cevaplamada itina göstermesidir. Ölçeğin maddelerine başına buyruk cevap veren, dalgınlık sonucu yanlış kategoriye onaylayan ya da araştırmacıların katılımcıların yanıtlarını puanlarken yapacakları hatalar ölçümün kalitesini olumsuz etkileyeceği kuşkusuzdur.

Rasch analizi çerçevesinde bir ölçümden elde edilen puanların kalitesi hakkında, madde ve kişi uyum istatistikleri ile yorumda bulunmak mümkündür. Uyum kavramsal olarak kişilerin ve maddelerin Rasch modelinden ne derece saptığının bir ölçüsüdür. Boone vd. (2014) yanıtlarında tutarsızlıklar bulunan birkaç bireyin varlığının maddelerin uyumlarına kayda değer seviyede ve olumsuz anlamda etkilediğini belirtmektedirler. Ölçeğin güvenilirlik ve geçerliği hakkında doğru yorumda bulunabilmek için analiz öncesinde verdiği yanıtlarda aşırı tutarsızlık bulunan kişilerin varlığına ilişkin bir inceleme gerçekleştirmiştir. Boone vd. (2014) önerileri doğrultusunda, Rasch modeline en az uyum gösteren maddelere bireylerin verdikleri yanıtları ve bu yanıtlar içerisinde en tutarsız olan bireyleri gösteren tablo (Winsteps, Tablo 11.1) incelenmiştir. Bu tablo içerisinde verdikleri yanıtların Z-artık (Z-residual) değerleri -3'ten küçük veya eşit ya da 3'ten büyük veya eşit olan bireylerin ilgili maddeye verdikleri yanıtlar silinmiştir. Diğer bir ifadeyle, bu kişilere ilgili maddeye hiç yanıt vermemiş gibi muamele edilmiştir. Rasch analizinde bir kişinin ölçüsünü belirlemek için kişinin testteki maddelerin tümüne cevap vermiş olması elzem olmadığından bu müdahale meşrudur (Boone ve Staver, 2020).

2.5. Araştırmanın etik izni

Yapılan bu çalışmada "Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesi" kapsamında uyulması belirtilen tüm kurallara uyulmuştur. Yönergenin ikinci bölümü olan "Bilimsel Araştırma ve Yayın Etiğine Aykırı Eylemler" başlığı altında belirtilen eylemlerden hiçbiri gerçekleştirilmemiştir.

Etik kurul izin bilgileri

Etik değerlendirmeyi yapan kurul adı: Trabzon Üniversitesi Sosyal ve Beşerî Bilimler Araştırma ve Etik Kurulu

Etik değerlendirme kararının tarihi: 06.01.2021

Etik değerlendirme belgesi sayı numarası: 81614018-000-E.20

3. BULGULAR

3.1. Testte Benimsenen 5'li Kategori Yapısının Uygunluğuna Yönelik Bulgular

Derecelendirme ölçekleri bir test maddesi için doğru/yanlış veya katılıyorum/katılmıyorum gibi iki kategorili yanıt seçeneklerinin sunabileceği bilgiden daha fazlasını elde etmek için kullanılmaktadır. Lakin, yanıt için iki kategori yerine, daha fazla kategorinin kullanılması bir maddeye yönelik güvenilir biçimde daha fazla bilgiye ulaşıldığını tek başına garanti etmemektedir. Daha fazla kategori kullanılarak istenilen amaca ulaşılp ulaşılmadığı, kategorilerin araştırmacının zihninde kurguladığı test ile örtük değişken arasındaki ilişkiye yönelik kavramsallaştırmayı resmetmede başarılı olup olmadığının incelenmesi gerekmektedir. Kategorilerin beklenen yönde işlediğinin bir kabul olarak ele alınması doğru bir yaklaşım olmayacaktır. Çünkü Andrich (1996) uygulamada testi cevaplayan kişilerin test geliştiricilerinin kurguladıklarından farklı bir biçimde kategorilere tepki gösterebildiğini belirtmektedir. Sonuç olarak bir testin geçerliğin arttıran kriterlerden birinin testin amaçlanan şekilde ölçme işlemini gerçekleştirmesi olduğu dikkate alındığında, bu inceleme geçerliğin sınanmasında atılması gereken bir adımdır. Testte kullanılan derecelendirme kategorilerinin işlevselliği derecelendirme ölçeği modeli ile incelenmiştir (Wright ve Masters, 1982).

İncelenen testin uygulandığı formda her bir madde için katılımcılara yanıtlayabilecekleri 5 kategori sunulmuştu (Ek-1). Bu kategoriler soldan sağa doğru ardışık olarak 1 ile 5 arasındaki tamsayılar ile puanlanmıştı. Dolayısıyla, görece sağ tarafta yer alan bir kategori solunda yer alan kategoriye nazaran daha üst seviyede öz yeterliğe hitap edeceği kurgulanmıştı. Bu kurgunun doğal sonucu, bir maddenin iki kategorisi için görece sağda bulunan onaylayan bireylerin ölçülerinin ortalaması, diğer deyişle öz yeterlik puanları ortalaması, diğerini onaylayanların ölçülerinin ortalamasından büyük olmasıdır.

Elde edilen verilerle yapılan ilk analizde 1. kategorinin (Hiçbir zaman) testin genelinde bireyler tarafından çok az onaylandığı (<%1) ayrıca bazı maddeler için hiç tercih edilmediği ortaya çıkmıştır. Bunun sonucu olarak araştırmacıların kategoriler için benimsedikleri kurgunun bazı maddeler için gerçekleşmediği görülmüştür. Örneğin, Tablo 3'te SS11-om kodlu madde için elde edilen veriler görülmektedir. Tablodan görüldüğü üzere 2. kategori için ortalama ölçü değeri 1. kategori için ortaya çıkan değerden daha düşüktür. Bir başka ifadeyle bu maddede 1. kategoriyi onaylayan bireylerin öz yeterlik puanları ortalaması, 2. kategoriyi onaylayan bireylerin öz yeterlik puanları ortalamasından büyüktür. Bu ise araştırmacıların kurgusu ile taban tabana zıttır.

Tablo 3

Ss11-Om Kodlu Madde İçin Kategorilerin Ortalama Ölçüleri

Madde	Kategori	Kategori değeri	Ortalama ölçü
SS11-om	Hiçbir zaman	1	0.74
	Ender olarak	2	0.45*
	Bazen	3	1.05
	Çoğu zaman	4	2.29
	Her zaman	5	4.58

1. kategorinin amaca hizmet etmediği tespiti sonrasında Linacre'nin (2002) önerileri temel alınarak 1. ve 2. kategoriler birleştirilip tek bir kategori haline getirilmiş, böylece kategori sayısı 4'e indirilmiştir. Ayrıca katılımcıların yanıtları tekrar puanlanmış (1-2-3-4) ve yeni puanlar üzerinde tekrardan Rasch analizi yapılmıştır. Bu çalışmada takip eden tüm analizler yeni puanlamadan elde edilen veriler üzerinde gerçekleştirilmiştir.

Linacre (2002) bir derecelendirme ölçeğinde yer alan kategorilerin işlevselliğini test etmek için şu kriterleri ileri sürmüştür; i) Her bir kategorinin frekansı en az 10 olmalıdır, ii) Kategori ölçüleri monoton artan olmalıdır, iii) Dış uyum istatistikleri 2'den küçük olmalıdır, iv) Kategori eşikleri monoton artan olmalıdır, v) Ardışık kategori eşikleri arasındaki mesafe 1.4 ile 5 logit arasında olmalıdır, vi) Kategori olasılık eğrilerinin her biri zirve yapmalıdır.

Kategorilerin yapısına yönelik ilk 5 kritere ait değerler Tablo 4'te sunulmuştur. Tablo 4'ten görüleceği üzere kategorilerin yapısına yönelik ileri sürülen ilk 5 kriter sağlanmaktadır.

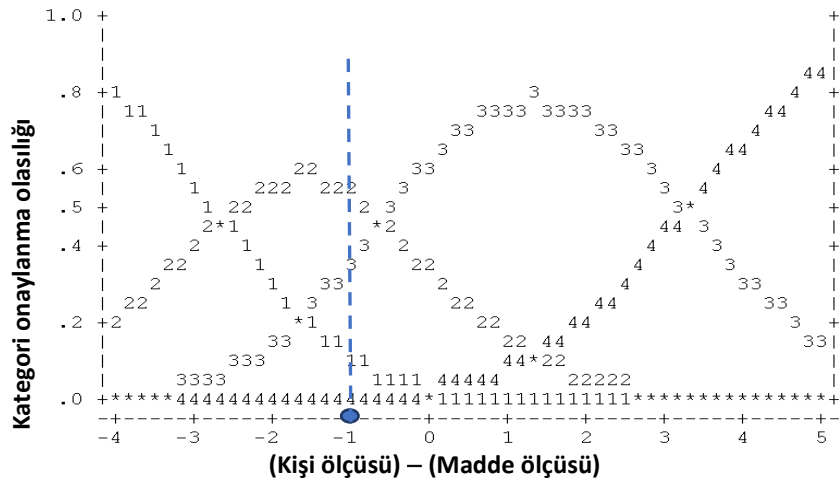
Tablo 4

Kategori Yapısına Yönelik İlk 5 Krite Ait Değerler

Kategoriler	Frekans (%)	Ortalama ölçü	Dış uyum (MnSq)	Kategori eşik (Andrich)
1. (Hiçbir zaman + Ender olarak)	260 (6)	-1.48	1.23	-
2. Bazen	1209 (27)	-0.65	0.87	-2.65
3. Çoğu zaman	2582 (57)	0.89	0.91	-0.62
4. Her zaman	478 (10)	2.25	1.06	3.27

Son kriteri sınamak için kategori olasılık eğrileri (Winsteps, Tablo 3.2+) oluşturulmuştur. Elde edilen olasılık eğrileri Şekil 2'de görülmektedir. Şekil 2'de yer alan grafikte düşey eksen olasılık değerlerini, yatay eksen ise kişi ölçüsü ile madde ölçüsü arasındaki farkı temsil etmektedir. Kategori olasılık eğrileri grafiği bir kişinin bir madde için bir kategoriyi onaylama olasılığını, kişinin ölçüsü ile maddenin ölçüsü arasındaki

farka dayalı olarak resmetmektedir. Somut bir örnek üzerinden açıklanacak olursa, A kişinin ölçüsü 1 logit ve B maddesinin ölçüsü 2 logit olsun. Bu durumda kişi ölçüsü ile madde ölçüsü arasındaki fark -1 logit olacaktır. Kategori olasılık eğrileri grafiğine bakıldığında bu kişinin ilgili madde için 4 numaralı kategoriye onaylama olasılığı sıfıra çok yakın, 1 numaralı kategoriye onaylama olasılığı yaklaşık %10, 3 numaralı kategoriye onaylama olasılığı yaklaşık %35 ve 2 numaralı kategoriye onaylama olasılığı yaklaşık %55 olmaktadır. Ayrıca, grafiklerin kesişme noktaları iki kategorinin eş olasılıkla seçilmesi anlamına gelmektedir. Bu noktaların apsisi Tablo 4'te Kategori eşik (Andrich) sütunundaki değerlere karşılık gelmektedir. Örneğin, bir kişinin ölçüsü ile bir maddenin ölçüsü arasındaki fark -2.65 logit ise bu kişinin 1 ve 2 numaralı kategorileri onaylama olasılığının eşit olduğu anlaşılmaktadır. Grafiğe bakıldığında her bir kategorinin onaylanma olasılığının en büyük olduğu değer bulunmaktadır. Linacre'nin (2002) ifadesiyle her bir kategori eğrisinin bir zirvesi mevcuttur. Böylece testin 4'lü kategori yapısı Linacre'nin ifade ettiği son kriteri de sağlamaktadır.

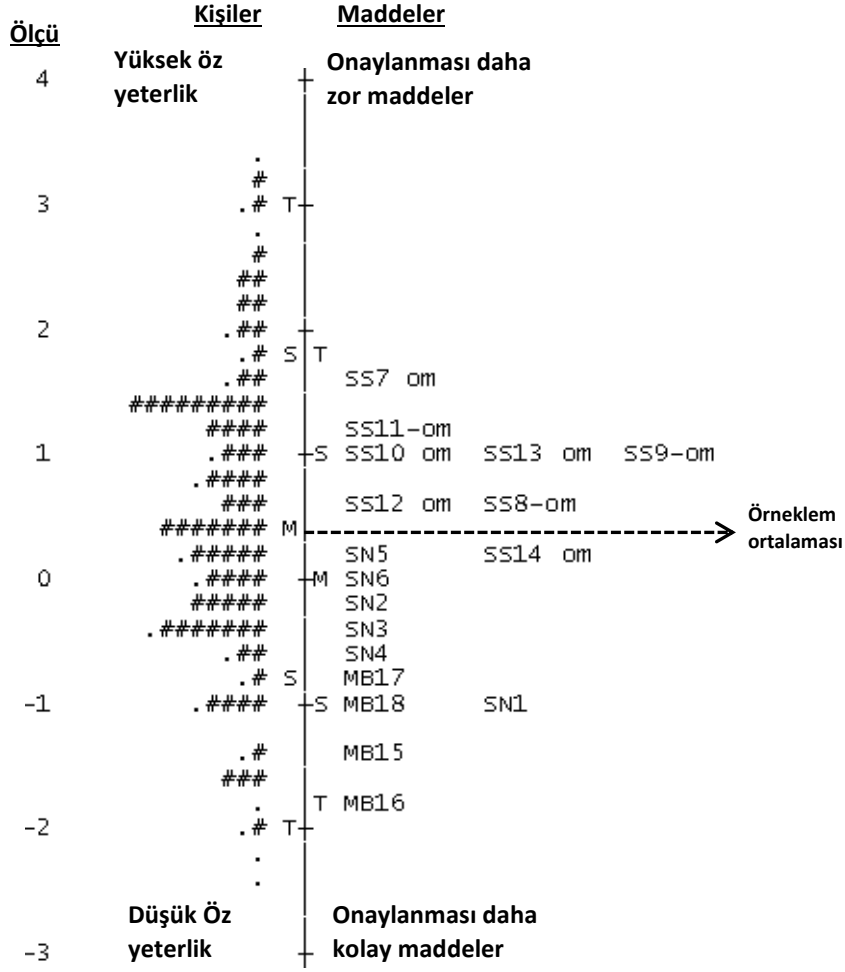


Şekil 2: Testin 4'lü kategori yapısına ait kategori olasılık eğriler

3.2. Madde Hiyerarşisi ve Testin Örnekleme Hedefleme Yeterliliğine İlişkin Bulgular

Bir testin kapsam geçerliliğini artıran faktörlerden biri test içerisinde testin ölçmeyi amaçladığı niteliğin farklı seviyelerine karşılık gelen maddelerin yer almasıdır (Boone vd., 2014). Eğer bir test içerisinde yer alan maddelerin güçlükleri dar bir aralık içerisinde hapsolmüş ise maddeler ölçülecek niteliğin düşük, orta ve yüksek seviyelerini ortaya çıkartmada yetersiz kalacaktır. Rasch analizi içerisinde test maddelerinin hiyerarşisi ve dolayısıyla ölçülecek niteliğin farklı seviyelerine hitap edip etmediğine yönelik çıkarım yapmada kullanılan araçlardan biri kişi-madde (Wright map) haritasıdır. Kişi-madde haritası teste tabi olan kişileri ve test içerisinde yer alan maddeleri birimi logit olan eş aralıklı ölçek üzerinde konumlandırmaktadır. Kişi-madde haritasının bir diğer işlevi ise testin uygulandığı örnekleme hedefleme yeterliliğine yönelik değerlendirme yapmayı mümkün kılmasıdır (Boone vd., 2014). Şekil 3'te bu çalışmada elde edilen kişi-madde haritası görülmektedir. Şeklin ortasında yer alan düşey doğru parçası kişilerin ve maddelerin ölçülerinin resmedildiği eş aralıklı ölçeği temsil etmektedir. Şeklin sol tarafında örnekleme yer alan bireylerin ölçülerinin dağılımı yer almaktadır. Buradaki "#" sembolü 3 kişiyi, "." sembolü ise 1 veya 2 kişiyi temsil etmektedir. Şeklin solunda ve sağında bulunan M, S ve T harfleri ise kişi ve madde ölçülerine ilişkin sırayla ortalama, 1 standart sapma ve 2 standart sapma değerlerini göstermektedir. Şeklin Kişiler kısmında aşağıdan yukarıya doğru gidildikçe katılımcıların öz yeterlik ölçüleri artmaktadır. Bir başka ifadeyle, yukarı doğru gidildikçe testten elde edilen ham puan artmaktadır. Sağ tarafta aşağıdan yukarı gidildikçe ise maddelerin onaylanması zorlaşmaktadır. Bir başka ifadeyle maddeye verilen

yanıtların karşılık geldiği toplam puan azalmaktadır. Bu ise daha yukarıda bulunan maddelerin öz yeterlik niteliğinin daha yüksek seviyelerini ölçtüğü anlamına gelmektedir.



Şekil 3: Kişi-madde haritası

Maddelerin konumları dikkate alındığında dar bir alanda kümelenmedikleri, dolayısıyla öz yeterlik niteliğinin farklı seviyelerine hitap ettiği görülmektedir. Bununla birlikte maddelerin arasında çok fazla boşluk bulunmamaktadır. Maddelerin arasında büyük boşlukların bulunması testin ilgili niteliği ölçme hususunda hassas olmadığı şeklinde değerlendirilmektedir (Boone vd., 2014). Maddeler arasında büyük boşlukların yer alması, bir kitabın ölçülerini belirlemek için birimleri cm yerine dm olan bir cetvelin kullanılmasına benzetilebilir. Haritada kişilerin dağılımına bakıldığında dar bir bölgede kümelenmedikleri görülmektedir. Bu durum test maddelerinin kişileri ilgili nitelik açısından ayırt etmede başarı gösterdiğine işaret etmektedir. Ayrıca test içerisinde onaylanması en zor olan SS7-om kodlu maddenin işaret ettiği seviyenin üzerinde kayda değer sayıda birey yer almaktadır. Ortaya çıkan bu dağılım, üzerinde çalışılan örneklem için test içerisinde onaylanması daha zor olan maddelerin bulunmasının, öz yeterliği yüksek olan bireyleri ayırt etmede testin daha iyi işlev göstereceğini belirtmektedir.

Bir testin uygulandığı örnekleme iyi hedeflemesi beklenir. Testin örnekleme iyi hedeflemesi, test içerisinde yer alan maddelerin örneklem için ne çok zor ne de çok kolay olması anlamına gelmektedir (Boone vd., 2014). Bir testin uygulandığı örnekleme iyi hedeflediğinin göstergelerinden biri maddelerin kümelenmediği bölge ile kişilerin kümelenmediği bölgenin karşılıklı olmasıdır (Bond ve Fox, 2007). Haritada kişi ve madde dağılımları karşılaştırıldığında her iki kısımda da yoğun bölgenin karşılıklı olduğu değerlendirilmektedir.

Testin örnekleme iyi hedefleyip hedeflemediğine karar kılmak için başvurulan diğer bir kriter örneklemin ortalaması ile maddelerin ortalamasının kıyaslanmasıdır. Rasch analizinde ortalama madde güçlüğü 0 logit değerine kalibre edilmektedir. Eğer kişilerin ortalaması yüksek pozitif bir değer ise bu testin örnekleme için kolay, düşük negatif bir değer ise testin örnekleme için zor olduğuna işaret etmektedir. Örnekleme ortalamasının 0.5 olması hafif düzeyde isabetsizliğe, 1 olması ise mühim bir isabetsizliğe işaret etmektedir (Duncan vd., 2003). Bu çalışmada kişi ölçüleri ortalaması 0.49 logit olarak ortaya çıkmıştır. Dolayısıyla test maddelerinin üzerinde çalışılan örnekleme iyi hedeflediği değerlendirilmektedir.

3.3. Madde Uyum İstatistiklerine İlişkin Bulgular

Kavramsal olarak uyum elde edilen verilerin Rasch modeline ne düzeyde riayet ettiği. Uyum kavramı araştırmacılara verilerin Rasch modelinin beklentisinden ne kadar saptığını belirlemede yardımcı olmaktadır. Eğer bir testte Rasch modeline uyum göstermeyen maddeler mevcutsa bu maddeler testte ölçülmesi hedeflenen değişkene hitap etmemektedir (Oon ve Subramaniam, 2011). Rasch analizi hem kişilere hem de test maddelerine yönelik uyum istatistikleri oluşturmaktadır. Bununla birlikte Wright ve Linacre (1994), madde ve kişi uyumu birlikte ele alındığında madde uyumunun kişi uyumundan daha ön planda olduğunu, dolayısıyla madde uyumunu gösteren istatistikler üzerinde daha fazla düşünülmesi gerektiğini ifade etmektedirler. Bunun gerekçesini ise aşağıdaki gibi açıklamaktadırlar.

İstatistiksel olarak Rasch açısından ele alındığında kişiler ve maddeler birebir aynıdır. Her ikisi de safi Rasch parametreleridir. Bu açıdan uyum kriterleri aynıdır. Lakin temelde maddeler ve kişiler farklılaşmaktadır. Maddelerin kişilere nazaran daha iyi davranmasını umut ederiz. Ayrıca madde güçlüklerinin gelecekte de devamlılık göstermesini, kişi yeterliliklerinin ise değişmesini bekleriz. Sonuç olarak, uyum açısından maddelere karşı tavrımız kişilere nazaran daha katıdır. Katılımcılar arasında birkaç başına buyruk kişinin mevcut olması bizi çok da endişelendirmez. Lakin birkaç başına buyruk maddenin mevcudiyeti testin uygulanması, verinin doğru bir şekilde girilip girilmediği, testte ele alınan örtük değişkenin tanımlanması vb. birçok hususta şüpheler ortaya çıkartır (s. 370).

Rasch analizi sonucunda elde edilen uyum istatistikleri artık kareler ortalaması (Mean-square residuals "MnSq") ile raporlaştırılmaktadır (Boone vd., 2014). Teknik açıdan, iç uyum (Infit) ve dış uyum (Outfit) birer Ki-kare istatistiği olup, verilerin ve modelin arasındaki ilişkinin düzeyini göstermektedir (Boone vd., 2014). Bu istatistiklerin beklenen değeri 1 olup sıfırdan büyük her değeri alabilmektedir (Wright ve Linacre, 1994). 1'den büyük olan MnSq değerleri verilerin modelin beklediğinden daha az öngörülebilir olduğuna, 1'den küçük değerler ise verilerin modelin beklediğinden daha çok öngörülebilir olduğuna işaret etmektedir. Örneğin, bir uyum indeksi değerinin 1.28 olması elde edilen verilerde Rasch modelinin öngördüğünden %28 daha fazla varyansın bulunduğu, .75 olması ise modelin öngördüğünden %25 daha az varyansın bulunduğu anlamına gelmektedir (Bond, Yan ve Heene, 2021). Dış uyum uçdeğerlere daha hassas olup, iç uyum uçdeğerleri daha az dikkate alır ve bir maddenin ölçüsüne yakın kişilerin yanıtlarına odaklanır (Boone vd., 2014). İç uyum ve dış uyum istatistikleri hususunda Wright ve Linacre (1994) tarafından farklı sınav türlerine yönelik artık kareler ortalaması için önerilen makul aralıklar Tablo 5'te görülmektedir.

Tablo 5

Farklı Sınav Türlerine İlişkin Artık Kareler Ortalaması Değeri İçin Önerilen Aralıklar

Sınav türü	Dış uyum (MnSq)
Katılımcılar için önemli sonuçları olan sınavlar	0.8 – 1.2
Alelade sınavlar	0.7 – 1.3
Derecelendirme Ölçeği	0.6 – 1.4
Klinik gözlem	0.5 – 1.7
Hakem değerlendirmesi	0.4 – 1.2

Uyum hususunda dikkate alınması önerilen bir diğer gösterge maddelerin nokta-ölçü korelasyonudur (point-measure correlation) (Boone ve Staver, 2020). Nokta-ölçü korelasyonu verideki gözlemler ile maddelerin ölçüleri arasındaki korelasyondur (Linacre, 2002). Bir maddenin nokta-ölçü korelasyon değerinin negatif olması o maddeye verilen yanıtların, maddelerin tümünün birlikte tanımladığı örtük değişken ile ters düştüğüne işaret etmektedir. Gerçekleştirilen Rasch analizi sonucunda maddelerin ölçüleri, uyum istatistikleri ve nokta-ölçü korelasyon değerleri (Winsteps, Tablo 13.1) Tablo 6'da sunulmuştur.

Tablo 6

Madde Uyum İstatistikleri

Madde	Madde ölçüsü (logit)	Dışuyum (MnSq)	İçuyum (MnSq)	Nokta-ölçü korelasyonu
SS7-om	1.57	1.06	1.02	.57
SS11-om	1.25	.96	.92	.66
SS9-om	1.01	1.09	1.05	.60
SS10-om	.98	.96	.99	.59
SS13-om	.94	.94	.91	.61
SS8-om	.54	1.07	1.05	.58
SS12-om	.51	1.03	1.05	.60
SS14-om	.17	1.12	1.14	.61
SN5	.11	.88	.89	.53
SN6	.08	.89	.92	.45
SN2	-.28	.87	.89	.50
SN3	-.32	.94	.93	.48
SN4	-.70	.68	.69	.54
MB17	-.77	.97	.97	.59
SN1	-.91	.69	.70	.55
MB18	-.95	1.12	1.11	.53
MB15	-1.44	1.37	1.36	.43
MB16	-1.80	1.27	1.30	.55

Tablo 6'da görüldüğü üzere maddelerin tamamı uyum istatistikleri önerilen aralık içerisinde dağılmakta olup nokta-ölçü korelasyon değerleri pozitifdir.

3.4. Güvenirlilik ve Ayırt edicilik İndekslerine İlişkin Bulgular

Rasch analizi ile hem kişiler hem de test maddelerinin güvenilirliği hakkında bilgi edinilebilmektedir. Kişi güvenilirliği klasik test teorisinde kullanılan güvenilirlik katsayıları (Cronbach alfa) ile benzer biçimde yorumlanmaktadır (Linacre, 2002). Dolayısıyla 1'e yakın değerler iç tutarlılığı yüksek bir ölçüme işaret eder. Rasch analizi testte yer alan maddeler için de güvenilirlik katsayısı oluşturmaktadır. Madde güvenilirliği, ölçülen değişken boyunca maddelerin belirgin bir hiyerarşi tanımlama kudreti olarak yorumlanabilir. Katsayının düşük olması testte yer alan maddelerin örtük değişkenin dar bir aralığını ölçtüğünü ifade eder (Linacre, 2002). Bir analoji ile örneklendirecek olursak, elimizde bir tartı olduğunu fakat bu tartının 60 ile 75 kg arasındaki ağırlıkları ölçtüğünü düşünelim. Böyle bir tartı ile bir mahalledeki kişilerin ağırlıklarını ölçmek istediğimizde yalnızca ağırlık değişkeninin dar bir aralığına (60-75 kg) düşen kişilerin ağırlıklarını belirleyebiliriz. Bir başka ifadeyle testte yer alan maddeler örtük değişkenin yüksek düzeyi ile düşük düzeyi arasında dağılmıyor demektir.

Yapılan ölçümde ham puanların güvenilirlik indeksi (Cronbach alfa) .86, Rasch ölçüleri temelinde kişi güvenilirlik indeksi .85, madde güvenilirlik indeksi ise .98 olarak hesaplanmıştır. Elde edilen kişi ve madde güvenilirlik indeksleri kabul edilebilirlik sınırı olan .7'den yüksektir (Boone vd., 2014).

Ayırt edicilik indeksi (G) kavramsal olarak 2 grup arasında gerçekleştirilen *t*-testine benzerdir. Bu indeks gerçek puanların standart sapmasının hataların standart sapmasına oranını vermektedir (Linacre, 2022). Kişi ayırt edicilik indeksi örneklemdaki bireyleri sınıflandırmak için kullanılmaktadır. Düşük ayırt edicilik indeksi ölçme aracının yüksek ve düşük performans sergileyen kişileri ayırt etmede yeterince hassas olmadığını göstermektedir. Kişi ayırt edicilik indeksinin 1.5 olması kabul edilebilir, 2 olması iyi ve 3 ya da daha yüksek olması mükemmel olarak nitelendirilmektedir (Duncan vd., 2003). Madde ayırt edicilik indeksinin ise birey bazında analiz yapabilmek için minimum 1.5, grup bazında analiz yapmak için ise minimum 2.5 olması gerekmektedir (Boone vd., 2014).

Bir analiz sonucunda örneklemdaki bireylerin kaç farklı tabakaya ayrılabilceği kişi ayırt edicilik indeksi kullanılarak aşağıdaki formülle belirlenmektedir (Linacre, 2022).

$$tabaka = \frac{4G + 1}{3}$$

Bu çalışmada gerçekleştirilen analiz sonucunda kişi ayırt edicilik indeksi 2.34, madde ayırt edicilik indeksi ise 7.84 olarak ortaya çıkmıştır. Yukarıdaki formülde kişi ayırt edicilik indeksi değeri yerine yazıldığında kişilere yönelik *tabaka* değeri yaklaşık olarak 3.5, maddelere yönelik ise yaklaşık 10.8 elde edilmektedir.

3.5. Testin Boyut Yapısına İlişkin Bulgular

Rasch analizinin temel gereksinimi olan tek boyutluluk kavramı, bir testi oluşturan maddelerde ortaya çıkan performansın tek bir örtük değişken tarafından açıklanabilirliği olarak tanımlanmaktadır (Brentari ve Silvia, 2007). Rasch teorisi çerçevesinde bir testin tek boyutlu olup olmadığını incelemeye kullanılan yöntem PCAR (Principal Component Analysis of Residuals) olarak isimlendirilen analizdir (Boone ve Staver, 2020). Rasch analizinde artıklar (residuals), gözlenen puanlar ile puanların beklenen değerleri arasındaki farktır (Wright ve Masters, 1990). PCAR, Rasch analizi sonunda hesaplanan puanların, veri seti kümesinde birden fazla boyutun varlığına ilişkin kanıt olup olmadığını tahlil eder ve olası kanıtların gücünü belirler (Boone ve Staver, 2020). PCAR, toplam varyanstan Rasch boyutunun ortaya çıkardığı varyansı çıkartarak, rastgeleliğin ve diğer olası etkilerin birlikte oluşturduğu artıkların varyansını ayırır. Rasch analizi sonucunda ortaya çıkan artıklar üzerinde gerçekleştirilen faktör analizinin amacı, artıkların varyansını en çok açıklayan ortak faktörün, böyle bir faktörün var olduğu varsayımı altında, ortaya çıkarılmasıdır (Linacre, 1998). Bir başka ifadeyle PCAR, verinin içerisinde Rasch modeli ile uyumlu olmayan desenlerin varlığını araştırır. Eğer maddelerin bir kısmının teşkil ettiği bir grup (cluster) aynı beklenmeyen desene sahipse o zaman bu grubun içerisindeki maddeler önemli ortak bir özelliği paylaşıyor olması muhtemeldir (Linacre, 2022). Eğer ortaya çıkarılan faktör sadece verideki rasgele gürültüden ileri geldiği belirlenirse, artıklar bağlamında anlamlı bir faktörün olmadığı kararına varılır (Linacre, 1998).

Eğer artıkların korelasyon matrisinden elde edilen birinci temel bileşenin özdeğeri (eigenvalue) 2'den küçük ise verideki gürültünün rasgele olduğu şeklinde yorumlanır (Boone ve Staver, 2020). Ayrıca Rasch analizi sonucunda elde edilen puanların işaret ettiği boyut toplam varyansın %50'den fazlasını açıklıyorsa bu da testin tek boyutlu olduğunun bir diğer göstergesidir (Boone ve Staver, 2020).

PCAR işlemi sonucunda maddeler 3 farklı kümeye ayrıştırılır. Bu kümeler, birbirleriyle yüksek korelasyon gösteren maddelerden oluşmaktadır. Kümeler arasındaki korelasyon değerinin 1'e yakın olması farklı kümelerde yer alan maddelerin aynı yapıyı ölçtüğüne işaret etmektedir (Linacre, 2022). Gerçekleştirilen PCAR analizi sonucunda ortaya çıkan kümeler ve bu kümelerin arasındaki korelasyon ve azaltılmamış korelasyon (dissattenuated correlation) değerleri Tablo 7'de sunulmuştur. Burada yer alan azaltılmamış korelasyon değeri, ölçümün hatasız olması durumunda korelasyon değerini ifade etmektedir. Tablo 7'de 1. ve 3. kümelerde yer alan maddelere bakıldığında dikkat çekici olan, olumlu ve olumsuz maddelerin ayrı ayrı kümelendikleridir.

Tablo 7

PCAR Sonucunda Ortaya Çıkan Madde Kümeleri ve Kümeler Arasındaki Korelasyon Değerleri

Madde Kümeleri			Kümeler arası korelasyon değerleri					
1	2	3	1-2	1-3	2-3			
SS9-om, SS8-om, SS11-om, SS13-om, SS14-om, SS10-om, SS12-om	SS7-om, SN6, SN3	MB15, MB17, MB16, MB18, SN4, SN5, SN2, SN1	Pearson Azaltılmamış	Pearson Azaltılmamış	Pearson Azaltılmamış	Pearson Azaltılmamış	Pearson Azaltılmamış	
			.5	.8	.35	.45	.48	.79

Artıklar üzerinde gerçekleştirilen temel bileşen analizi sonucunda Rasch ölçülerinin açıkladığı varyansın oranı %43, birinci bileşenin öz değeri ise 3.2 olarak ortaya çıkmıştır. Bu değerler sırasıyla %50'den küçük ve 2'den büyüktür. Dolayısıyla bu durum, 1. ve 3. kümelerde bulunan maddelerin farklı boyutları ölçüyor olabileceği ihtimalini ortaya çıkarmaktadır. Böyle bir durumda Linacre (2022), 1. ve 3. kümeler arasındaki azaltılmamış korelasyon katsayısı .7'den büyükse iki kümedeki maddelerin aynı yapıyı ölçtüğünün, .3'ten küçükse farklı yapıyı ölçtüğünün çıkarsanabileceğini belirtmektedir. Tablo 6'dan görüldüğü üzere 1. ve 3. kümeler arasındaki azaltılmamış korelasyon katsayısı (.45) bu iki değer arasındadır. Dolayısıyla bu iki kümedeki maddelerin aynı yapıyı ölçüp ölçmediği hususunda şu ana kadar belirtilen kriterler ile çıkarım yapmak mümkün görünmemektedir.

Linacre (2022), ortaya çıkan bu belirsizliğin giderilmesi için elde edilen ölçümler temelinde oluşturulacak simülasyon veri seti üzerinde PCAR analizinin gerçekleştirilmesini önermektedir. Simüle edilmiş veri setinde, gözlenen verilerin analizi sonucunda elde edilen kişi ölçüleri, madde güçlükleri ve derecelendirme ölçeğinin yapısı sabit kalmaktadır. Değiştirilen ise kişilerin bazı maddelere verdikleri yanıtlardır. Bu değişim, eğer ölçüm yapılırken hiç hata yapılmamış olsaydı bu kişilerin hangi kategorileri onaylayacağını tahmini üzerine dayanmaktadır. Bir diğer ifadeyle simüle edilmiş veri seti, Rasch modeline mükemmel uyum gerçekleşmesi durumunda elde edilecek veri setinin resmini sunmaktadır. Bu öneriden hareketle simülasyon veri seti oluşturularak PCAR analizi gerçekleştirilmiştir. Analizin sonucunda birinci temel bileşenin öz değeri 1.57 ve madde kümeleri arasındaki azaltılmamış korelasyon değerleri .75 (1-3), .99 (1-2), .99 (2-3) olarak elde edilmiştir. Sonuç olarak, istatistikî yöntemler temelinde testte yer alan maddelerin aynı özelliği ölçtüğü savını çürütecek yeterli delil bulunmamaktadır.

3.6. Madde Yanlılığına İlişkin Bulgular

Bir testin yapı geçerliliğini etkileyen unsurlardan biri, testte yer alan maddelerin örneklemin alt kümelerinin her birinde yanlı davranıp davranmadığıdır. Bu durumu sınamak için gerçekleştirilen Değişen Madde Fonksiyonu Analizi (Differential Item Functioning [DMFA]) bir testin içerisinde yer alan maddelerin örneklemin içerisinde yer alan farklı gruplar için aynı şekilde işlev gösterip göstermediğini analiz etmektedir. Rasch modeli bağlamında bir maddenin işlevindeki farklılaşma, testin ölçtüğü niteliğe eşit düzeyde sahip olan fakat farklı gruplarda bulunan bireylerin bir maddeye yanıt verme olasılıklarının anlamlı derecede farklı olması durumunda ortaya çıkmaktadır (Linacre, 2022). Bir başka ifadeyle, maddelerin güçlük açısından sıralamasının örneklemin alt gruplarında farklılaşmasıdır (Boone vd., 2014). Bu çalışmada örneklemin oluşturulan bireyleri farklılaştıran en önemli 3 değişken cinsiyet, öğrenim gördükleri kurum ve sınıf düzeyidir. Testteki maddelerin bu değişkenlerin belirlediği alt kümelerde benzer işlev gösterip göstermediğini belirlemek için her bir değişken için Winsteps programı vasıtasıyla DMFA gerçekleştirilmiştir. Boone vd., (2014) iki grup arasında bir maddenin yanlılık gösterdiğine karar kılmak için madde ölçüleri arasındaki farkın mutlak değerinin .64'ten büyük ve olasılık değerinin .05'ten küçük olması gerektiğini kriter olarak önermektedir.

Testte yer alan maddelerin işlevinin cinsiyet değişkeni bağlamında yanlılık gösterip göstermediğini belirlemek için gerçekleştirilen analiz sonucunda elde edilen bulgular Tablo 8’de sunulmuştur. Gerçekleştirilen analizde her bir madde için “Madde güçlükleri iki grupta eşittir” hipotezi test edilmiştir (Winsteps, Tablo 30.1). Tablo 8’den görüldüğü üzere tüm maddelerin iki gruptaki ölçümlerinin farkının mutlak değeri kriter olarak sunulan .64’ten düşük ve olasılık değerleri de .05’ten büyüktür.

Tablo 8
Cinsiyet Değişkeni Açısından Gerçekleştirilen DMFA Bulguları

Madde	Kadın		Erkek		DMF fark (Kadın-Erkek)	t	p
	DMF ölçüsü	Std. hata	DMF ölçüsü	Std. hata			
MB15	-1.50	.14	-1.16	.30	-.34	-1.04	.2213
SS12-om	.51	.12	.51	.27	.00	.00	.4865
SS10-om	1.02	.12	.76	.26	.26	.92	.3197
SS7-om	1.63	.11	1.30	.26	.34	1.20	.2630
SN6	.12	.12	-.13	.28	.25	.82	.3167
SN3	-.37	.13	-.05	.28	-.32	-1.05	.2812
SN2	-.24	.13	-.45	.29	.21	.67	.6533
MB16	-1.80	.14	-1.80	.31	.00	.00	.8221
SS14-om	.11	.12	.48	.27	-.37	-1.24	.5897
SN5	.08	.12	.26	.27	-.18	-.59	.6525
SN4	-.66	.13	-.89	.30	.22	.68	.7768
SN1	-.95	.13	-.71	.29	-.24	-.74	.2025
SS13-om	.94	.12	1.03	.26	-.09	-.31	.7704
MB17	-.77	.13	-.71	.29	-.06	-.19	.2438
MB18	-.93	.13	-1.07	.30	.14	.41	.7292
SS9-om	1.01	.12	1.01	.26	.00	.00	.8419
SS8-om	.57	.12	.40	.27	.17	.58	.2258
SS11-om	1.25	.12	1.30	.26	-.05	-.18	.5665

Test maddelerinin işlevselliğinin örneklemdaki bireylerin öğrenim gördükleri üniversite açısından farklılaşma durumunu belirlemeye yönelik gerçekleştirilen DMFA’dan elde edilen bulgular Tablo 9’da verilmiştir. Tablo 9’da bulunan Ü1, Marmara bölgesindeki bir devlet üniversitesini; Ü2, Karadeniz bölgesindeki bir devlet üniversitesini temsil etmektedir. Gerçekleştirilen analizde “Madde güçlükleri iki grupta eşittir” hipotezi test edilmiştir (Winsteps, Tablo 30.1). Tablo 9’dan görüldüğü üzere MB 15 kodlu madde için elde edilen olasılık değeri anlamlılık düzeyinden düşüktür. Bu maddenin iki gruptaki ölçümleri arasındaki fark ise .56 olup Boone, vd.’in (2014) önerdiği değerden biraz düşüktür. İlgili madde Ü2 üniversitesinde öğrenim görmekte olan öğrenciler için Ü1’de öğrenim görmekte olan öğrencilere nazaran onaylanması daha kolay bir maddedir.

Tablo 9

Öğrenim Görülen Kurum Değişkeni Açısından Gerçekleştirilen DMFA Bulguları

Madde	Ü1		Ü2		DMF fark (Ü1-Ü2)	t	p
	DMF ölçüsü	Std. hata	DMF ölçüsü	Std. hata			
MB15	-1.07	.22	-1.63	.15	.56	2.09	.0382*
SS12-om	.23	.20	.63	.13	-.40	1.64	.1018
SS10-om	1.02	.19	.96	.13	.07	.29	.7732
SS7-om	1.44	.19	1.64	.13	-.20	-.87	.3835
SN6	-.01	.20	.13	.14	-.14	-.58	.5660
SN3	-.27	.21	-.20	.14	-.37	1.47	.1441
SN2	-.31	.21	-.28	.14	-.03	-.14	.8914
MB16	-1.95	.22	-1.74	.15	-.21	-.77	.4410
SS14-om	.43	.20	.05	.14	.38	1.56	.1197
SN5	-.14	.21	.22	.14	-.36	1.46	.1457
SN4	-.7	.21	-.70	.15	.00	.00	1.000
SN1	-.98	.22	-.87	.15	-.11	-.43	.6645
SS13-om	1.16	.19	.84	.13	.33	1.43	.1554
MB17	-.80	.21	-.77	.15	-.03	-.11	.9157
MB18	-.80	.21	-1.03	.15	.23	.87	.3880
SS9-om	1.27	.19	.89	.13	.38	1.67	.0961
SS8-om	.47	.20	.58	.13	-.11	-.48	.6293
SS11-om	1.27	.19	1.25	.13	.02	.11	.9164

Testte yer alan maddelerin sınıf seviyesi bağlamında işlevselliğini incelemek için gerçekleştirilen DMFA sonucunda elde edilen veriler Tablo 10'da görülmektedir. Linacre'nin (2022) önerisi doğrultusunda, gerçekleştirilen analizde her bir madde ve her sınıf düzeyi için "Maddenin örneklemden hesaplanan güçlüğü ile gruptan hesaplanan güçlüğü eşittir" hipotezi test edilmiştir (Winsteps, Tablo 30.2). Tablo 10'dan görüldüğü üzere hiçbir madde ve sınıf seviyesi için elde edilen olasılık değerleri anlamlılık sınırından küçük değildir. Bununla birlikte SN3 kodlu madde için 1. sınıflarda elde edilen değer anlamlılık düzeyine yakın çıkmıştır. İlgili maddenin farklı sınıf seviyelerindeki DMF ölçüleri dikkate alındığında maddenin 1. sınıf öğrencileri için diğer sınıf seviyelerindeki öğrencilere nispeten onaylanması daha zor bir madde olduğu görülmektedir.

Tablo 10

Sınıf Seviyesi Değişkeni Açısından Gerçekleştirilen DMFA Bulguları

Madde	Sınıflar												
	1. Sınıf				2. Sınıf			3. sınıf			4. Sınıf		
	Örnekleme ölçüsü	DMF ölçüsü	t	p	DMF ölçüsü	t	p	DMF ölçüsü	t	p	DMF ölçüsü	t	p
MB15	-1.44	-1.36	.40	.68	-1.59	-.68	.50	-.92	1.69	.10	-2.05	-1.65	.11
SS12-om	.51	.51	.00	.99	.42	-.45	.66	.46	-.15	.87	.84	1.03	.31
SS10-om	.98	.71	-1.57	.12	1.21	1.23	.22	.86	-.43	.66	1.43	1.49	.15
SS7-om	1.57	1.57	.00	.99	1.67	.53	.59	1.22	-1.31	.20	1.79	.73	.47
SN6	.08	.34	1.47	.14	-.18	-1.29	.20	-.10	-.62	.53	.10	.06	.95
SS14-om	.17	.12	-.28	.78	.07	-.52	.60	.17	.00	.99	.65	1.47	.15
SN5	.11	.25	.79	.43	.07	-.20	.83	-.01	-.42	.67	-.13	-.69	.50
SN4	-.70	-.81	-.59	.56	-.81	-.51	.60	-.54	.52	.61	-.25	1.32	.20

Tablo 10 Devamı

Madde	Örneklem ölçüsü	Sınıflar											
		1. Sınıf				2. Sınıf			3. sınıf			4. Sınıf	
		DMF ölçüsü	t	p	DMF ölçüsü	t	p	DMF ölçüsü	t	p	DMF ölçüsü	t	p
MB18	-.95	-1.25	-1.55	.13	-.76	.86	.39	-.73	.72	.47	-.74	.61	.55
SS9-om	1.01	1.15	.86	.39	1.07	.29	.76	.63	-1.37	.18	.85	-.50	.62
SS8-om	.54	.37	-1.03	.30	.66	.58	.56	.48	-.24	.80	.95	1.30	.20
SS11-om	1.25	1.25	.00	.99	1.28	.19	.85	1.22	-.08	.94	1.25	.00	.99
SN3	-.32	.03	1.94	.06	-.49	-.82	.41	-.54	-.75	.46	-.86	-1.53	.14
SN2	-.28	-.23	.27	.78	-.36	-.39	.70	-.10	.60	.55	-.49	-.60	.55
MB16	-1.80	-1.90	-.48	.63	-1.75	.25	.80	-1.61	.62	.54	-1.92	-.30	.76
SN1	-.91	-.88	.12	.90	-.86	.22	.82	-.54	1.20	.23	-1.65	-2.02	.06
SS13-om	.94	.91	-.21	.83	.94	.00	.99	.94	.00	.99	1.05	.35	.73
MB17	-.77	-.77	.00	.99	-.67	.47	.64	-.54	.76	.46	-1.38	-1.66	.11

3.7. Yerel Bağımsızlık Kriterine İlişkin Bulgular

Rasch teorisinin varsayımlarından biri, bir niteliği ölçmek için kullanılacak testte yer alan maddelerin yerel bağımsızlık göstermesidir (Bond vd., 2021). Yerel bağımsızlık, bir maddeye verilen yanıtların başka bir maddeye verilen yanıtlardan etkilenmemesi anlamına gelmektedir (Fan ve Bond, 2019). Eğer maddeler birbirlerinden bağımsızlarsa, veriden Rasch boyutu çıkartıldıktan sonra madde artıklarının birbirleriyle yüksek korelasyon sergilememesi gerekmektedir (Fan ve Bond, 2019). Linacre (2022), iki maddenin yerel bağımlı olduğuna şüphelenmek için artıkları arasındaki korelasyon katsayısının .7'den büyük olmasını kriter olarak sunmaktadır. Bu çalışmada gerçekleştirilen analiz sonucunda artıklar arasındaki korelasyon katsayıları -.41 ile .39 arasında değerler almıştır. Buradan hareketle ölçekte yer alan maddelerin yerel bağımsızlık kriterini sağladığı değerlendirilmektedir.

4. TARTIŞMA, SONUÇ ve ÖNERİLER

Araştırma kapsamında cevap aranan problemlerden biri SSBKÖYÖ maddelerini cevaplama benimsenen 5 kategorili yapının işlevsel olup olmadığını belirlemektir. Elde edilen veriler benimsenen kategori yapısının araştırmacıların ölçek hazırlama sürecinde kurguladıkları işleve tam olarak hizmet etmediğini göstermiştir. Bu olumsuz durumun ortaya çıkmasına sebep olan ana etmen, ilk yanıt kategorisi olan "Hiçbir zaman" seçeneğinin testin genelinde öğrencilerin çok az bir kısmı tarafından onaylanması, bazı maddeler içinse hiç onaylanmaması olmuştur. Birinci ve ikinci kategorilerin birleştirilerek tek bir kategori haline getirilmesi kategori yapısını daha işlevsel hale getirmiştir. Oluşturulan yeni 4 kategorili yapı literatürde kategori işlevselliği için önerilerin kriterlerin tamamını sağlamıştır. Ölçeğin uygulandığı örneklemin ilköğretim matematik öğretmeni adayları olduğu dikkate alındığında, birinci kategori için ortaya çıkan durumun esasında şaşırtıcı olmadığı değerlendirilmektedir. Çünkü, örneklemdaki bireylerin okudukları program sebebiyle matematiğe karşı genel kitleye nazaran daha yetenekli ve matematiksel bilgiye sahip olma açısından daha üst düzeyde olduğu söylenebilir. Bu açıdan ele alındığında, ölçek içerisinde yer alan maddelerin birçoğunda bu kategoriyi onaylamamaları aslında beklenen bir durumdur. Bu sebepten, matematik öğretmeni adayları üzerinden gerçekleştirilecek ileriki çalışmalarda ilk kategori haricindeki kategorilerin kullanılması önerilmektedir. Bununla birlikte, ortaöğretim öğrencileri gibi matematiksel

bilgiye sahip olma hususunda daha fazla deęişkenlik gösterebilecek örneklem grubunda ilk kategorinin işlevsel olabileceęi düşünölmektedir. Bu husus ileriki arařtırmalarda cevap bulacak bir husustur.

Arařtırmada ele alınan problemlerden bir dięeri testin madde hiyerarřının uygunluęuna ve örnekleme hedeflemede yeterlilięine iliřkindi. Testte yer verilen maddelerin hiyerarřisi oluşturulan kiři-madde haritası ile resmedilmiřtir. Kiři-madde haritasında maddeler açasından ortaya çıkan desen dikkate alındığında maddelerin öz yeterlik seviyesini temsil eden doęru boyunca geniř bir aralıkta daęılım gösterdięi ortaya çıkmıřtır. Bu durum testte öz yeterlik nitelięinin farklı seviyelerine karřılık gelen maddelerin bulunduęu anlamına gelmektedir. Dięer taraftan maddeler arasında büyük boşluklar bulunmamaktadır. Bu ise testte yer alan maddelerin bir bütün olarak öz yeterlik nitelięini ölçmede hassaslıęa sahip olduęuna işaret etmektedir. Sonuç olarak, ölçęin öz yeterlik nitelięini geniř bir aralıęını kabul edilebilir hassaslıkta ölçtüęü deęerlendirilmektedir.

Bir ölçęin yapı geçerlięi ile iliřkili hususlardan biri, ölçek maddeleri için belirlenen puanların, bir dięer anlamda onaylanma zorluęunun, arařtırmacıların beklentilerine ve ortak akla uygun olup olmadıęıdır (Boone ve Staver, 2020). Test maddelerini oluştururken arařtırmacıların deneyimlerine dayanarak temel aldıkları varsayım; genel olarak matematiksel ifadeleri oluřturma ve anlam çıkartmanın, ifadeler içerisindeki bileřenleri tanımaya nazaran daha zor olacaęı yönündeydi. Arařtırmacıları bu varsayıma iten gözlemlerinden biri, öğrencilerin sınavlardan elde ettikleri puanlara tepki olarak dile getirdikleri “Hocam biliyorum ama yazamıyorum” tarzındaki serzenişlere defaatle tanık olmalarıdır. Bir dięer gözlem ise öğrencilerin sembolik biçimde ifade edilmiş önermeleri okumakta ve anlamını açıklamada genel olarak sıkıntı yaşamalarıdır. Arařtırmacıların öğretim sürecinde deneyimledikleri bu olgulara benzer sonuçlar literatürde yer alan farklı arařtırmalarda da (Akyıldız ve Çınar, 2016; Çakmak vd., 2014; Yeřildere, 2007) rapor edilmiřtir. Kiři-madde haritasına bakıldıęında onaylanması en zor olan madde SS7-om olup, madde bir matematiksel ifadenin sembolik biçimini yazma öz yeterlilięine iliřkindir. Peři sıra gelen ilk 4 maddeye (SS11-om, SS10-om, SS13-om, SS9-om) bakıldıęında matematiksel ifadelerin sembolik biçimin oluřturma ve onlardan anlam çıkartma ile ilgilidir. Dięer taraftan onaylanması en kolay ilk 4’ü MB16, MB15, MB18 ve SN1 maddeleridir. Bu maddelerin içerięine bakıldıęında ise kullanılan sembol ve baęlaçların doęrudan tercümesine iliřkin olduęu anlaşılmaktadır. Sonuç olarak, yapılan ölçüm maddelerin onaylanma zorluklarını arařtırmacıların deneyimleri ve ortak akla paralel olarak ortaya çıkarmıř olup, madde hiyerarřisi bu anlamda tutarlıdır.

Ölçekte yer alan maddeler bir bütün olarak örneklem için onaylanması ne çok zor ne de çok kolaydır. Örneklem ortalamasının pozitif olması ölçęin örneklem için bir nebze kolay olduęunu belirtmektedir. Fakat bu fark literatürde öne sürölen deęerden daha küçüktür. Kiři-madde haritasında ise maddeler ve bireyler karřılıklı konumlanmıřtır. Tüm bu göstergelerden ölçęin örnekleme iyi hedefledięi deęerlendirilmektedir.

Ölçekte yer alan maddelerin tamamının uyum indeksleri literatürde önerilen aralık içerisindeydir. Dolayısıyla ölçek maddeleri Rasch modeli ile uyum sergilemiřtir. Bu durum testteki maddelerin bir bütün olarak ölçölen nitelięe hitap ettięini göstermektedir (Oon ve Subramaniam, 2011). Dięer yandan tüm maddelerin nokta-ölçü korelasyon deęerleri pozitifdir. Sonuç olarak, bu göstergeler ölçęin her maddesinin testin ölçmeyi hedefledięi nitelięe hizmet ettięine işaret etmektedir.

Ham puanlar üzerinden elde edilen güvenilirlik indeksi deęeri ölçümün yüksek düzeyde iç tutarlık sergiledięini ortaya koymuřtur. Rasch analizi sonucunda elde edilen kiři ve madde güvenilirlik indeksleri ise literatürde önerilen eřik deęerlerden daha yüksektir. Sonuç olarak, maddelerin onaylanma zorluklarının farklı bir örneklem üzerinde yapılacak ölçümde alacaęı deęerlerin bu çalışmada elde edilen deęerlere yakın olması, başka bir ifadeyle kiři-madde haritasında konumlarının aynı kalması yüksek ihtimaldir (Bond vd., 2021). Kiři ve madde güvenilirliklerine baęlı olarak elde edilen ayırt edicilik indeksleri de literatürde önerilen eřik deęerlerden yüksektir. Kiři ayırt edicilik indeksine baęlı olarak bulunan tabaka deęeri 3’ten büyüktür. Bu deęer, ölçekte yer alan maddelerin kiřileri 3 farklı tabakaya ayırt edebildięine işaret etmektedir. Sonuç olarak, ölçek örnekleme bireyleri düşük öz yeterlik, orta seviye öz yeterlik, yüksek öz yeterlik olarak isimlendirilebilecek 3 farklı gruba ayırma potansiyeline haizdir. Dięer taraftan maddelere yönelik ortaya

çıkan yüksek tabaka değeri, maddelerin ölçekte hedeflenen niteliği ölçmede geniş bir aralığı kapsadığı sonucunu ortaya koymaktadır.

Testin boyut yapısını incelemek için gerçekleştirilen ilk PCAR analizi, testin Rasch boyutundan farklı bir boyut daha barındırdığına yönelik şüphe oluşturmuştur. Bunun akabinde, böyle bir durumla karşılaşılması halinde literatürde atılması tavsiye edilen adımlar takip edilmiş ve sonuçta ölçeğin Rasch boyutundan farklı bir boyut içerdiği şüphesi istatistikî anlamda doğrulanamamıştır. Rasch analizi çerçevesinde bir testin boyut yapısına ilişkin karar vermede nicel göstergeler kesin belirleyici değildir. Bunlardan daha önemlisi, araştırmacıların maddelere yönelik nitel değerlendirmelerinin sonunda sunacağı gerekçeler ile varacağı kanaattir (Boone ve Staver, 2020). Bir başka ifadeyle, sadece öz değer ve Rasch boyutunun açıkladığı varyansın kriter alınarak testin Rasch boyutundan farklı bir boyut daha içerdiği sonucuna varılması sağlıklı bir değerlendirme olmayacaktır. Bu hususta Linacre'nin (2022) getirdiği açıklamalardan biri aşağıdaki gibidir.

Bir matematik testinde cebirsel problemler ile sözel problemler arasında büyük bir zıtlık elde ederiz. Bu iki problem türü kavramsal olarak farklıdır lakin her ikisi de matematiğin bir parçasıdır. Bu iki gruptaki maddeleri testten atmak istemeyiz ve ayırmak da istemeyiz. Yapabileceğimiz en iyi şey, testi oluştururken kurguladığımız plan ile uyumlu olacak şekilde bu iki gruptaki maddelerin sayılarını dengelemek olacaktır (s. 627).

Yine bu hususta Linacre'nin (2022) sunduğu bir başka izah aşağıdaki gibidir.

Özetle, zıtlık oluşturan maddelerin ifade ediliş biçimine bakın. Eğer bu maddeler farklı boyutlara hitap eder nitelikte ise ("ağırlık" ve "uzunluk" örneklerinde olduğu gibi) farklı boyutları oluşturan maddelerin analizlerini ayrı ayrı gerçekleştirin. Eğer maddeler aynı boyutun farklı kısımlarına ait ise (bir aritmetik testinde "toplama" ve "çıkarma" alanlarına ait olan maddeler gibi) eyleme gerek yoktur. Böyle bir durumda görülen, bir boyutun farklı kısımlarına hitap eden maddelerin beklenen kovaryansıdır (s. 627).

Araştırmacıların ölçek maddelerinin içeriğine yönelik yaptıkları nitel değerlendirme sonucunda vardıkları karar maddelerin tamamının matematik dilini kullanabilme öz yeterliliği ile ilgili olduğudur. Dolayısıyla test maddelerinin bir bütün halinde bu öz yeterliliği ölçtüğü değerlendirilmektedir. Lakin şüpheye sebep olan durumda ilginç olan, 1 numaralı kümedeki maddelerin tamamının olumsuz, 3 numaralı kümede yer alan maddelerin ise olumlu olmasıdır. Karşılaşılan bu durumun aynısı Oon ve Subramaniam'ın (2011) çalışmasında da ortaya çıkmış ve araştırmacılar analiz ettikleri ölçeğin gelecekte uygulanacağı zaman olumsuz maddelerin olumlu biçime dönüştürülmesini önermişlerdir.

Likert tipi ölçeklerde katılımcıların maddeleri cevaplarırken monotonlaşmasını önlemek, dikkatlerini canlı tutmak ve tek bir yönde yanıt verme yanlılığına engel olmak için olumlu ve olumsuz maddelere yer verilmektedir (Boone ve Staver, 2020; Zeng, Wen ve Zhang, 2020). Lakin, Carlson vd. (2011) tarafından gerçekleştirilen ve olumsuz maddelerin psikometrik özelliklerinin incelenmesini amaçlayan araştırmada, olumsuz maddelerin olumlulara nazaran daha az iç tutarlık ve toplam test puanları ile daha zayıf ilişki sergiledikleri rapor edilmiştir. Benzer amaçla yürütülen bir başka araştırmada Suárez-Alvarez vd. (2018), olumlu ve olumsuz maddelerin bir arada kullanılması durumunda test güvenilirliğinin olumsuz etkilendiğini, puanların varyansının azaldığını ve testin tek boyutluluk özelliğinin ikincil varyans kaynakları sebebiyle riske girdiğini belirtmektedir. Yine bir başka çalışmada Zeng vd. (2020), olumlu ve olumsuz maddelerin bir arada kullanılmasının testin iç tutarlılığını zayıflattığını ve ölçeğin faktör yapısını değiştirdiğini belirtmektedir.

Olumsuz maddelere yönelik literatür dikkate alındığında, öz yeterlik ölçeğinin tek boyutlu olduğu ve PCAR analizinin işaret ettiği ek boyutun olumsuz maddelerin ölçekte yer almasından kaynaklandığı değerlendirilmektedir. Boone vd., (2014) kendi deneyimleri ışığında ölçekler de olumsuz madde

kullanmamayı önermektedir. Bu çalışmada ortaya çıkan durum bu araştırmacıların deneyimlerini doğrulamaktadır. Sonuç olarak, gelecekte ölçeğin uygulanmasında olumsuz maddelerin ifadelerinin değiştirilerek olumlu forma dönüştürülmesi önerilmektedir. Tablo 11’de araştırmacıların bu dönüşüm için önerisi yer almaktadır.

Tablo 11
Ölçeğin Olumsuz Maddelerine Karşılık Önerilen Olumlu Maddeler

Mevcut madde kodu	Mevcut madde	Önerilen madde
SS7-om	Sözel biçimde verilen bir matematiksel ifadeyi, bir kaynaktan yararlanma ihtiyacı duymadan sembolik biçimde yazmakta zorlanırım.	Sözel biçimde verilen bir matematiksel ifadeyi, bir kaynaktan yararlanma ihtiyacı duymadan sembolik biçimde yazabilirim.
SS8-om	Sözel biçimini anladığım bir matematiksel ifadenin sembolik biçimini anlamakta zorlanırım.	Sözel biçimini anladığım bir matematiksel ifadenin sembolik biçimini de anlayabilirim.
SS9-om	Sembolik biçimde ifade edilen bir teoremin ne iddia ettiğini anlamakta zorlanırım.	Sembolik biçimde ifade edilen bir teoremin ne iddia ettiğini anlayabilirim.
SS10-om	Sınavlarda zihnimdeki düşünceleri sembolik biçimde yazarken sembolleri doğru şekilde kullanmakta zorlanırım.	Sınavlarda zihnimdeki düşünceleri sembolik biçimde yazarken sembolleri doğru şekilde kullanabilirim.
SS11-om	Sembolik biçimdeki matematiksel ifadeleri anlatırken akıcı ve açık bir dil kullanmada zorlanırım.	Sembolik biçimdeki matematiksel ifadeleri anlatırken akıcı ve açık bir dil kullanabilirim.
SS12-om	Sınavlarda çözüm yaparken matematiksel sembolleri doğru olarak kullanmakta zorlanırım.	Sınavlarda çözüm yaparken matematiksel sembolleri doğru olarak kullanabilirim.
SS13-om	Yalnızca sembolik biçim kullanılarak oluşturulan bir matematiksel ifadeyi sözel olarak ifade etmekte zorlanırım.	Yalnızca sembolik biçim kullanılarak oluşturulan bir matematiksel ifadeyi sözel olarak ifade edebilirim.
SS14-om	Sembolik biçimini bildiğim bir matematiksel ifadenin sözel biçimini anlamakta zorlanırım.	Sembolik biçimini bildiğim bir matematiksel ifadenin sözel biçimini de anlayabilirim.

Ölçekteki maddelerin yanlılık gösterip göstermediğini belirleme amacıyla 3 değişken temelinde DMFA gerçekleştirilmiştir. Cinsiyet değişkeni açısından hiçbir madde yanlılık emaresi göstermemiştir. Coğrafi bölge açısından ise MB15 kodlu madde anlamlılık sınırı yakın bir olasılık değeri ortaya çıkarmıştır. Lakin ilgili madde her iki grup için de onaylanması en kolay olan 2’nci maddedir. Diğer bir ifadeyle tüm maddeler arasında maddenin güçlük sırası her iki grupta da aynıdır. Sonuç olarak maddenin yanlı davranmadığı değerlendirilmektedir.

Testteki maddelerin sınıf seviyesi değişkeni açısından yanlılıklarını tahlil etmek için yapılan DMFA analizinde elde edilen bulgular, SN3 kodlu maddenin 1. sınıf öğrencileri için diğer sınıf seviyelerindeki öğrencilere nazaran onaylanması daha zor olduğunu göstermiş ve elde edilen olasılık anlamlılık düzeyine yakın çıkmıştır. İlgili madde, öğrencilerin matematiksel notasyonların hangi matematiksel kavramları temsil ettiğini bilmeye ilişkin öz yeterliklerini ölçmeye yöneliktir. Bu farklılığa, 1. sınıf öğrencilerinin üniversite düzeyindeki akademik matematikle yeni tanışıyor olması ve öğretim hayatlarında ilk kez bu kadar yoğun sembol içeren matematiksel ifadeler ile karşılaşılıyor olmasının sebep olduğu değerlendirilmektedir. Sonuç olarak, ele alınan değişkenler açısından testin madde yanlılığı göstermediği, ele alınan örneklem alt gruplarında madde hiyerarşisinin benzer olduğu değerlendirilmektedir.

Ölçek geliştirme uzun soluklu bir süreçtir. Bir ölçeğin ortaya konduğu ilk formunun mükemmel işlev göstermesi çok zor bir ihtimaldir (Boone ve Staver, 2020). Bu gerçekten hareketle, bu çalışmada ortaya konan önerilerin alanda test edilmesi gerekmektedir. Bilhassa olumsuz maddelere yapılacak dönüşümün ölçeğin

yapısında ve madde hiyerarşisinde değişiklik meydana getirip getirmediği ileriki çalışmalarda farklı örneklemeler üzerinde ele alınması önerilmektedir.

Kaynakça / Reference

- Açıl, E., & Zeybek, Z. (2017). Öğrencilerin matematiksel dili kullanma ve anlama becerisi ile öğretmenlerin öğrencilerin matematiksel dili nasıl kullandıklarını fark edebilme yeteneği. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 42(42), 87-107. <https://dergipark.org.tr/en/download/article-file/399675>
- Akyıldız, P., & Çınar, C. (2016). İlköğretim matematik öğretmen adaylarının lineer cebir dersine yönelik tutumları ve alan dili yeterliklerinin incelenmesi. *Türk Eğitim Bilimleri Dergisi*, 14(1), 1-22. <https://dergipark.org.tr/en/download/article-file/267659>
- Andrich, D. (1978). Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied Psychological Measurement*, 2(4), 581-594.
- Andrich, D. (1996). A hyperbolic cosine latent trait model for unfolding polytomous responses: Reconciling Thurstone and Likert methodologies. *British Journal of Mathematical and Statistical Psychology*, 49, 347-365. <https://doi.org/10.1111/j.2044-8317.1996.tb01093.x>
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191-215.
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). New Jersey: NJ: Lawrence Erlbaum Associates.
- Bond, T. G., Yan, Z., & Heene, M. (2021). *Applying the Rasch model: Fundamental measurement in the human sciences* (4th ed.). New York: NY: Routledge.
- Boone, W. J., & Staver, J. R. (2020). *Advances in Rasch analysis in the human sciences*. Cham, Switzerland: Springer. <https://doi.org/10.1007/978-3-030-43420-5>
- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Dordrecht, Netherlands: Springer. https://doi.org/10.1007/978-94-007-6857-4_4
- Brentari, E., & Silvia, G. (2007). Unidimensionality in the Rasch model: How to detect and interpret. *Statistica*, 67(3), 253-261.
- Büyükoztürk, Ş., Kılıç Çakmak, E., Akgün, Ö. E., Karadeniz, Ş., & Demirel, F. (2019). *Eğitimde bilimsel araştırma yöntemleri* (26. baskı). Pegem Akademi.
- Carlson, M., Wilcox, R., Chou, C.-P., Chang, M., Yang, F., Blanchard, J., . . . Clark, F. (2011). Psychometric properties of reverse-scored items on the CES-D in a sample of ethnically diverse older adults. *Psychol Assess*, 23(2), 582-562. <https://doi.org/10.1037/a0022484>
- Cevahir, R. (2022). *Öğretmen adaylarının matematiksel ifadelerin sembolik ve sözel biçimleri arasında tercüme yapabilme öz yeterliklerinin incelenmesi* (Yüksek lisans tezi). Yükseköğretim Kurulu Ulusal Tez Merkezi'nden edinilmiştir. (Tez No. 707742).
- Çakmak, Z., Bekdemir, M., & Baş, F. (2014). İlköğretim matematik öğretmenliği öğrencilerinin örüntüler konusundaki matematiksel dil becerileri. *Erzincan Üniversitesi Eğitim Fakültesi Dergisi*, 16(1), 204-223. <https://dergipark.org.tr/en/download/article-file/68233>
- Çekmez, E. (2020). Öğretmen adaylarının önermelerin sembolik ve sözel formları arasında tercüme yapabilme becerilerinin incelenmesi. *Trakya Eğitim Dergisi*, 10(2), 551-565. <https://doi.org/10.24315/tred.642192>
- Çepni, S. (2007). *Araştırma ve proje çalışmalarına giriş*. Trabzon: Celepler Matbaacılık.
- Demirci, F. (2021). *Bilişim teknolojileri öğretmenlerinin ve öğretmen adaylarının öz-yeterlik kaynakları ile teknolojik pedagojik alan bilgileri arasındaki ilişki* [Doktora tezi, Ankara Üniversitesi]. <https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp>
- Duncan, P. W., Bode, R. K., Lai, S. M., & Perera, S. (2003). Rasch analysis of a new stroke-specific outcome scale: the stroke impact scale. *Archives of Physical Medicine and Rehabilitation*, 84(7), 950-963. [https://doi.org/10.1016/S0003-9993\(03\)00035-2](https://doi.org/10.1016/S0003-9993(03)00035-2)
- Fan, J., & Bond T. (2019). Applying Rasch measurement in language assessment: Unidimensionality and local independence. In V. Aryadoust & M. Raquel (Eds.), *Quantitative data analysis for language assessment, Vol. I: Fundamental techniques* (pp.83-102). Routledge. <https://doi.org/10.4324/9781315187815>

- Fırat, Z. S., & Dinçer, Ç. (2018). Okul öncesi öğretmenlerinin doğal matematiksel dil kullanımlarının incelenmesi. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 18(2), 895-914. <https://dergipark.org.tr/en/download/article-file/427601>
- Kabael, T., & Baran, A. A. (2016). Matematik öğretmenlerinin matematiksel iletişim becerilerinin gelişimine yönelik farkındalıklarının incelenmesi. *Elementary Education Online*, 15(3), 868-881. <https://dergipark.org.tr/en/download/article-file/225930>
- Kabael, T., & Yayan, B. (2017). Effect of self-evaluation on pre-service mathematics teacher' self-efficacy in language of mathematics. *Anadolu Journal of Educational Sciences International*, 7(1), 1-34. <https://dergipark.org.tr/en/pub/ajesi/issue/27726/292575>
- Linacre, J. M. (1998). Structure in Rasch residuals: Why principal components analysis (PCA)? *Rasch Measurement Transactions*, 12(2), 636. <https://www.rasch.org/rmt/rmt122m.htm>
- Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3(1), 85-106.
- Linacre, J. M. (2022, 08 20). *A user guide to Winsteps Ministep Rasch model computer programs*. Winsteps: <https://www.winsteps.com/a/Winsteps-Manual.pdf> adresinden alındı
- Neumann, I., Neumann, K., & Nehm, R. (2011). Evaluating instrument quality in science education: Rasch-based analyses of nature of science test. *International Journal of Science Education*, 33(10), 1373-1405. doi:<https://doi.org/10.1080/09500693.2010.511297>
- Oon, P.-T., & Subramaniam, R. (2011). Rasch modelling of a scale that explores the take-up of physics among school students from the perspective of teachers. In R. F. Cavanagh, & R. F. Waugh, *Applications of Rasch Measurement in Learning Environments Research* (2nd ed., pp. 119-139). Rotterdam: Sense Publishers.
- Otterburn, M. K., & Nicholson, A. R. (1976). The language of (CSE) mathematics. *Mathematics in School*, 5(5), 18-20.
- Suárez-Alvarez, J., Pedrosa, I., Lozano, L. M., García-Cueto, E., Cuesta, M., & Muñiz, J. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema*, 30(2), 149-158. doi:<https://doi.org/10.7334/psicothema2018.33>
- Tschannen-Moran, M., & Hoy, A. W. (2007). The differential antecedents of self-efficacy beliefs and novice experienced teachers. *Teaching and Teacher Education*, 23(6), 944-956. <https://doi.org/10.1016/j.tate.2006.05.003>
- Tuchman, E., & Isaacs, J. (2011). The influence of formal and informal formative pre-service experiences on teacher self-efficacy. *Educational Psychology*, 31(4), 413-433. <https://doi.org/10.1080/01443410.2011.560656>
- Uğurel, I., & Moralı, S. (2010). Bir ortaöğretim matematik dersindeki ispat yapma etkinliğine yönelik sınıfçı tartışma sürecine öğrenci söylemleri çerçevesinde yakından bakış. *Dokuz Eylül Üniversitesi Buca Eğitim Fakültesi Dergisi*(28), 134-154. <https://dergipark.org.tr/en/download/article-file/231681>
- Uygur-Kabael, T. (2017). Ortaokul matematik öğretmen adaylarının informalden formal matematik diline çevirme becerileri: Türkiye ve Amerika Birleşik Devletleri durumları. *Hacettepe Eğitim Fakültesi Dergisi*, 32(4), 1013-1031.
- Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8(3), 370.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago: Mesa Press.
- Wright, B. D., & Masters, G. N. (1990). Computation of outfit and infit statistics. *Rasch Measurement Transactions*, 3(4), 84-85. <https://www.rasch.org/rmt/rmt34e.htm>
- Yeşildere, S. (2007). İlköğretim matematik öğretmen adaylarının matematiksel alan dilini kullanma yeterlikleri. *Boğaziçi Üniversitesi Eğitim Dergisi*, 24(2), 61-70. <https://dergipark.org.tr/tr/pub/buje/issue/3824/51415>
- Yıldırım, C. (2018). *Matematiksel düşünme* (14. b.). İstanbul: Remzi Kitabevi.

Zeng, B., Wen, H., & Zhang, J. (2020). How does the valence of wording affect features of a scale? The method effects in the undergraduate learning burnout scale. *Frontiers in Psychology, 11*, 26-38. <https://doi.org/10.3389/fpsyg.2020.585179>

EKLER

Ek-1. Sembolik ve sözel biçimi kullanma öz yeterlik ölçeği (SSBKÖYÖ)

SEMBOLİK VE SÖZEL BİÇİMİ KULLANMA ÖZ YETERLİK ÖLÇEĞİ						
#	MADDELER	1=HIÇBİR ZAMAN	2=ENDER OLARAK	3=BAZEN	4=ÇOĞU ZAMAN	5=HER ZAMAN
Sembol ve Notasyon Bilgisi						
SN1	Ders esnasında kullanılan sembollerin ne anlama geldiğini bilirim.					
SN2	Bir matematiksel kavramın sözel tanımına karşılık gelen sembolik tanımı tanıyabilirim.					
SN3	Matematiksel notasyonların hangi matematiksel kavramı temsil ettiğini bilirim.					
SN4	Matematiksel işlemlerde yer alan sembollerin anlamını doğru şekilde açıklayabilirim.					
SN5	Matematik derslerinde geçen kavramları hangi matematiksel notasyonlarla temsil edeceğimi bilirim.					
SN6	Sözel olarak ifade edilmiş teoremlerin sembolik biçimini tanıyabilirim.					
Sembolik ve Sözel Biçim İlişkilendirme						
SS7-om	Sözel biçimde verilen bir matematiksel ifadeyi, bir kaynaktan yararlanma ihtiyacı duymadan sembolik biçimde yazmakta zorlanırım.					
SS8-om	Sözel biçimini anladığım bir matematiksel ifadenin sembolik biçimini anlamakta zorlanırım.					
SS9-om	Sembolik biçimde ifade edilen bir teoremin ne iddia ettiğini anlamakta zorlanırım.					
SS10-om	Sınavlarda zihnimdeki düşünceleri sembolik biçimde yazarken sembollerini doğru şekilde kullanmakta zorlanırım.					
SS11-om	Sembolik biçimdeki matematiksel ifadeleri anlatırken akıcı ve açık bir dil kullanmada zorlanırım.					
SS12-om	Sınavlarda çözüm yaparken matematiksel sembollerini doğru olarak kullanmakta zorlanırım.					
SS13-om	Yalnızca sembolik biçim kullanılarak oluşturulan bir matematiksel ifadeyi sözel olarak ifade etmekte zorlanırım.					
SS14-om	Sembolik biçimini bildiğim bir matematiksel ifadenin sözel biçimini anlamakta zorlanırım.					
Matematiksel Bağlaç Kullanma						
MB15	Birden çok bağlaç ($\wedge, \vee, \Rightarrow, \Leftrightarrow, \forall, \dots$) içeren matematiksel ifadeleri doğru şekilde okuyabilirim.					
MB16	Sembolik biçimi verilen mantıksal bağlaçların ($\wedge, \vee, \Rightarrow, \Leftrightarrow, \forall, \dots$), sözel olarak nasıl ifade edildiğini bilirim.					
MB17	Sözel olarak verilen önermeleri uygun matematiksel bağlaçlar ($\wedge, \vee, \Rightarrow, \Leftrightarrow, \forall, \dots$) kullanarak yazabilirim.					
MB18	Sözel ifadelerde kullanılan mantıksal bağlaçlara karşılık gelen sembollerini bilirim.					

EXTENDED ABSTRACT

1. INTRODUCTION

Although it is emphasized in the curricula prepared by national and international official institutions that students should be able to use mathematics effectively as a means of communication, studies conducted at different levels of education (Çekmez, 2020; Otterburn and Nicholson, 1976; Uygur-Kabael, 2017) have reported that students' mathematical communication skills are not at a sufficient level and that students have difficulties in using mathematical language correctly. In order for students to be exposed to the language of mathematics, mathematics teachers need to use the language of mathematics correctly and emphasize the language of mathematics explicitly or implicitly in their lessons. One of the important factors that will enable mathematics teachers to use mathematical language correctly in teaching is their self-efficacy in using mathematical language. When the Turkish mathematics education literature was examined, no test was found to measure pre-service mathematics teachers' self-efficacy towards the language of mathematics. In order to contribute to this gap, a scale was developed by the second author with the Classical Test Theory approach in her master's thesis study completed under the supervision of the first author. In this study, it is aimed to get more detailed evidence about the validity and reliability of the related test by utilizing the advantages offered by Rasch Measurement Theory. In line with this purpose, the main problem addressed in the research is the extent to which the prepared scale performs valid and reliable measurement within the framework of the following questions.

- Is the category structure adopted for answering the scale items appropriate?
- Are individual and item reliability and discrimination indices appropriate?
- Are the fit statistics of the items within the acceptable range?
- Does the scale target the sample well?
- Is the hierarchy and distribution of items appropriate?
- Do the scale items show bias?
- Does the scale include a dimension different from the Rasch dimension?
- Does the scale meet the criterion of local independence?

2. METHOD

The participants of the study consisted of a total of 252 pre-service teachers studying in the elementary mathematics teaching programs of two education faculties located in the Marmara and Black Sea regions. In the study a scale which was developed by the second author was used as a data collection tool. The latent variable that the scale aims to measure is self-efficacy regarding the ability to translate between symbolic and verbal forms of mathematical expressions. There are 18 items in the scale, 8 of which are negative. The items were scored such that the higher the raw score obtained, the higher the self-efficacy quality. The scale sent to the prospective elementary mathematics teachers via an online form. The Rasch analysis was carried out with Winsteps® (version: 3.91) software.

3. FINDINGS, DISCUSSION AND RESULTS

The data obtained showed that the adopted category structure did not fully serve the function that the researchers had designed during the scale preparation process. The main factor that led to the emergence of this negative situation was that the first response category, "Never", was endorsed by a very small proportion of the students throughout the test and was not endorsed at all for some items. Combining the first and second categories into a single category made the category structure more functional. The new 4-category structure provided all of the criteria recommended in the literature for category functionality. For this reason, it is recommended to use categories other than the first category in future studies to be

conducted on prospective mathematics teachers. Another problem addressed in the study was the appropriateness of the item hierarchy of the test and its adequacy in targeting the sample. The hierarchy of the items in the test was depicted with the person-item map. Considering the pattern that emerged in terms of items in the person-item map, it was revealed that the items were distributed in a wide range along the line representing the self-efficacy level. As a result, the scale is considered to measure a wide range of self-efficacy quality with acceptable precision.

The fit indices of all items in the scale are within the range recommended in the literature. This shows that the items in the test address the measured quality. On the other hand, the point-measure correlation values of all items are positive. As a result, these indicators suggest that each item of the scale serves the quality that the test aims to measure. The reliability index value obtained over the raw scores revealed that the measurement exhibited a high level of internal consistency. The person and item reliability indices obtained are higher than the threshold values suggested in the literature. The discrimination indices obtained based on person and item reliabilities are also higher than the threshold values suggested in the literature. The stratum value found based on the person discrimination index is greater than 3. This value indicates that the items in the scale can discriminate individuals into 3 different strata. As a result, the scale has the potential to categorize the individuals in the sample into 3 different groups that can be named as low self-efficacy, moderate self-efficacy, and high self-efficacy.

Differential item functioning analysis was conducted based on 3 variables to determine whether the items in the scale showed bias. In terms of gender variable, no item showed any sign of bias. In terms of geographical region, the item coded MB15 revealed a probability value close to the significance limit. However, the related item was the second easiest item to approve for both groups. In other words, the order of difficulty of the item among all items is the same in both groups. As a result, it is evaluated that the item is not biased. The findings related to the grade level variable showed that the item coded as SN3 was more difficult to approve for 1st grade students than for students in other grade levels, and the probability obtained was close to the significance level. The related item aims to measure students' self-efficacy in knowing which mathematical concepts mathematical notations represent. This difference is attributed to the fact that first grade students are new to academic mathematics at the university level and that they are encountering mathematical expressions containing so many symbols for the first time in their education life. As a result, it was concluded that the test did not show item bias in terms of the variables considered and that the item hierarchy was similar in the sample subgroups.

ARAŞTIRMANIN ETİK İZİNİ

Bu çalışmada “Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesi” kapsamında uyulması gerektiği belirtilen tüm kurallara uyulmuştur. Yönergenin ikinci bölümü olan “Bilimsel Araştırma ve Yayın Etiğine Aykırı Eylemler” başlığı altında belirtilen eylemlerden hiçbiri gerçekleştirilmemiştir.

Etik kurul izin bilgileri

Etik değerlendirmeyi yapan kurul adı: Trabzon Üniversitesi Sosyal ve Beşerî Bilimler Araştırma ve Etik Kurulu

Etik değerlendirme kararının tarihi: 06.01.2021

Etik değerlendirme belgesi sayı numarası: 81614018-000-E.20

ARAŞTIRMACILARIN KATKI ORANI

Araştırmada 1. yazarın araştırmaya katkı oranı %60, 2. yazarın araştırmaya katkı oranı %40'tır. Bunun yanı sıra hangi araştırmacıların araştırmanın katkıda bulunduğu aşamalar aşağıda belirtilmiştir.

Yazar 1: Araştırmanın tasarlanması, Rasch teorisi bağlamında veri analizi, yorumlanması ve raporlaştırma, ölçeğin oluşturulmasında danışmanlık.

Yazar 2: Veri toplama, araştırmanın tasarlanması, matematik dili literatürünün raporlaştırılması.

ÇATIŞMA BEYANI

Araştırmada çıkar çatışması bulunmamaktadır.