

PAPER DETAILS

TITLE: İlkokul 3. ve 4. Sınıf Öğrencilerine Yönelik Matematik Kaygı Ölçeğinin Psikometrik Özelliklerinin İncelenmesi

AUTHORS: Erdem Çekmez,Gamze Tuti,Samet Yavuz Terzi

PAGES: 819-841

ORIGINAL PDF URL: <https://dergipark.org.tr/tr/download/article-file/2965604>



İlkokul 3. ve 4. Sınıf Öğrencilerine Yönelik Matematik Kaygı Ölçeğinin Psikometrik Özelliklerinin İncelenmesi

Investigation of the Psychometric Properties of a Mathematics Anxiety Test for 3rd and 4th Grade Primary School Students

Erdem ÇEKMEZ¹, Gamze TUTİ², Samet Yavuz TERZİ³

Makale Türü⁴: Araştırma Makalesi

Başvuru Tarihi: 21.02.2023

Kabul Tarihi: 19.08.2023

Atf İçin: Çekmez, E., Tuti, G. ve Terzi, S. Y. (2023). İlkokul 3. ve 4. sınıf öğrencilerine yönelik matematik kaygı ölçeğinin psikometrik özelliklerinin incelenmesi. *Anadolu Üniversitesi Eğitim Fakültesi Dergisi (AUJEF)*, 7(4), 819-841.

ÖZ: Bu araştırmanın amacı, ilkökulda öğrenim görmekte olan öğrencilerin matematik kaygı düzeylerini ölçmek için Mutlu ve Söylemez (2018) tarafından geliştirilmiş testin psikometrik özelliklerini Rasch Ölçme Teorisi çerçevesinde incelemektir. Bu amaçla test, madde uyum indeksleri, kategori yapısının işlevselliği, madde ve kişilere yönelik güvenirlik ve ayırt edicilik indeksleri, örneklemi hedefleme yeterliği, maddelerin dağılımı, cinsiyet ve sınıf düzeyi değişkenleri açısından test maddelerinin yanlılık gösterip göstermediği ve boyut yapısı bileşenleri bağlamında analiz edilmiştir. Araştırmanın katılımcılarını Trabzon ilinde farklı ilkökullarda öğrenim görmekte olan 1110 (581 kız, 529 erkek) üçüncü sınıf ve 946 (518 kız, 428 erkek) dördüncü sınıf olmak üzere toplamda 2056 öğrenci oluşturmaktadır. Test öğrencilere uygulandıktan sonra maddelere ve kişilere yönelik ölçüler Derecelendirme Ölçeği Modeli ile Winsteps (Sürüm: 3.91) yazılımı kullanılarak hesaplanmıştır. Elde edilen bulgular benimsenen kategori yapısının önerilen bazı kriterleri ihlal ettiğini, testin tek boyutluluk kriterini sağladığını, maddelerin ölçülen değişkenin farklı seviyelerini kapsadığını, kişi ve madde güvenirlik indekslerinin uygun olduğunu, maddelerin kişileri ayırt etmede düşük performans gösterdiği ve maddelerin cinsiyet ve sınıf düzeyi açısından yanlılık sergilemediğini göstermiştir. Bununla birlikte testin uygulandığı grubu hedeflemeye kayda değer bir isabetsizlik sergilediği ortaya çıkmıştır. Elde edilen bulgular temelinde test maddelerine ve maddeleri cevaplamak için benimsenen kategori yapısına ilişkin önerilerde bulunulmuştur.

Anahtar sözcükler: Matematik kaygısı, ilkökul düzeyi matematik eğitimi, Rasch ölçme teorisi

¹ Doç. Dr., Trabzon Üniversitesi, Fatih Eğitim Fakültesi, erdemcekmez@gmail.com, ORCID: 0000-0001-8684-2820 (Başlıca Yazar)

² Dr., Trabzon İl Millî Eğitim Müdürlüğü, gmzeelskn@gmail.com, ORCID: 0000-0001-8831-6613

³ Uzman Öğretmen, Trabzon İl Millî Eğitim Müdürlüğü, sametyavuzterzi@hotmail.com, ORCID: 0000-0002-4403-7759

⁴ Araştırma Trabzon Üniversitesi 2023-2/1.13 sayılı etik kurul izni ile yürütülmüştür.

ABSTRACT: The aim of this study is to investigate the psychometric properties of the mathematics anxiety test developed by Mutlu and Söylemez (2018) to measure the anxiety levels of 3rd and 4th grade students towards mathematics within the framework of Rasch Measurement Theory. To this aim, the test was analyzed in terms of item fit indices, functionality of the category structure, reliability and separation indexes of items and persons, sample targeting efficacy, distribution of items, unidimensionality, and whether the items show differential item functioning with respect to gender and grade level. The participants of the study consisted of 1110 (581 girls, 529 boys) third grade and 946 (518 girls, 428 boys) fourth grade, 2056 in total, students studying in different primary schools in Trabzon region. Upon the administration of the test, item calibrations and person measures were calculated by using the Rating Scale Model and Winsteps (version: 3. 91) software. The findings showed that the category structure adopted to respond the test items violated some of the suggested criteria in the literature, the test met the assumption of unidimensionality, the item and person reliability indexes were acceptable, the items covered a wide range of the latent variable, the items showed poor performance in separating persons, the items did not show differential item functioning in terms of gender and grade level variables, In addition to these, the test exhibited a remarkable inaccuracy in targeting the sample. Based on the findings, some suggestions were put forth regarding test items and the category structure adopted to respond to test items.

Keywords: Mathematics anxiety, primary mathematics education, Rasch measurement theory

1. GİRİŞ

Matematik öğrenmede bireyin sahip olduğu bilişsel özelliklerin yanı sıra matematiğe ve matematik öğrenme ortamlarında bulunan bileşenlere yönelik sahip olduğu duyuşsal nitelikler de önemli bir role sahiptir. Duyuşsal bağlamda matematik öğrenme sürecinde etkili olan faktörlerden biri matematik kaygısıdır. Manav (2011) kaygı kavramını genel anlamda, “herhangi bir tehlikenin korkusunun yansıması olarak insanda ortaya çıkan tedirginlik ya da akıl dışı korku” şeklinde tanımlamıştır (s. 202). Matematik öğrenme özelinde ise Bekdemir (2009) yurtdışı literatürde farklı araştırmacıların yaptıkları tanımları sentezleyerek matematik kaygısını “matematiğin kullanılması gerektiği durumlarda ortaya çıkan, belirtileri fiziksel olarak da gözüken, matematikle ilgili çözümü, öğrenme ve başarıyı, kısaca matematik yapmayı engelleyen mantık dışı panik, telaş, utanma, kaçınma, başaramama ve korku duygusu” olarak tanımlamıştır (s. 171).

İlkokulda öğrenim görmekte olan öğrencilerin matematik kaygısı yaşamayabileceği düşüncesinden ötürü bu öğretim seviyesinde yapılan araştırmalar diğer öğretim seviyelerine kıyasen sayıca daha azdır (Mutlu ve Söylemez, 2018). Lakin; Mutlu, Söylemez ve Yasul (2017) tarafından gerçekleştirilen araştırmada ilkokul 3. ve 4. sınıf seviyesinde öğrenim görmekte olan öğrencilerin yüksek düzeyde matematik kaygısına sahip olduğu belirlenmiştir. Yine ilköğretim seviyesinde gerçekleştirilen bir başka araştırmada Yenilmez ve Özbey (2006), alt sınıflarda öğrenim görmekte olan öğrencilerin üst sınıflarda öğrenim görmekte olan öğrencilere nispeten daha yüksek kaygı düzeyine sahip olduklarını belirlemişlerdir.

Öğrencilerde matematik dersinde gerekli becerileri sergileyemeyecekleri yönünde oluşan yanlış inanış, matematiğe yönelik kaygı düzeyinin artmasına ve bunun sonucu olarak matematikten giderek uzaklaşmalarına sebep olmaktadır (Türk ve Bedir, 2021). Bu yanlış inanış devamında gerçeğe dönüşmekte ve bunun sonucunda öğrenciler, matematik dersinde ele alınan içeriği anlayamamalarından ötürü başarısız olmaktadır (Yenilmez ve Özbey, 2006). Matematik başarısı ile matematik kaygısı arasındaki ilişkiyi odağa alan araştırmalar bu iki değişken arasında negatif yönlü bir ilişkinin olduğunu göstermektedir (Yenilmez ve Özbey, 2006). Bu konu çerçevesinde Ma (1999) matematik kaygısı ile matematik başarısı arasındaki ilişkiye yönelik gerçekleştirdiği araştırmasında 26 çalışmayı meta-analiz yöntemiyle incelemiştir. Elde ettiği bulgular matematik başarısı ile matematik kaygısı arasında negatif yönlü anlamlı bir ilişki olduğu sonucunu ortaya koymuştur. Bununla birlikte bu ilişki sınıf düzeyi, cinsiyet, etnik köken, kaygıyı ölçmek için kullanılan testler açısından tutarlılık sergilemiştir. Türk ve Bedir (2021) tarafından 4. sınıf düzeyinde öğrenim görmekte olan öğrenciler üzerinde yürütülmüş araştırmada da matematik kaygısı ile matematik başarısı arasında negatif yönlü anlamlı bir ilişki bulunmuştur.

Literatürde yer alan araştırmalardan anlaşıldığı üzere matematik kaygısı öğretim çağının her döneminde öğrencilerde ortaya çıkabilecek bir durumdur. Bununla birlikte matematik derslerinde yer verilen içeriğin anlaşılmasında da önemli bir role sahiptir. Dolayısıyla, matematik dersinde kaygıdan kaynaklı başarısızlıkların önüne geçebilmek için matematik kaygısını deneyimleyen öğrencilere gerekli desteğin verilmesi gerekmektedir. Bu desteğin verilebilmesi için matematik kaygısına sahip olan öğrencilerin tespiti gerçekleştirilmesi gereken ilk adımdır. Millî Eğitim Bakanlığının 2022 yılında başlattığı matematik seferberliği kapsamında Trabzon Millî Eğitim Müdürlüğü’nün gerçekleştirdiği matematik çalıştay çerçevesinde il bazında öğrencilerin matematik kaygı düzeylerini belirlenmesi amaçlanmıştır. Bu amaç doğrultusunda, 3. ve 4. sınıflarda öğrenim görmekte olan öğrenciler örneklem alınarak ve Mutlu ve Söylemez (2018) tarafından geliştirilen matematik kaygı ölçeği kullanılarak bir tarama gerçekleştirilmiştir. Bu çalışmada, yapılan ölçüm temelinde ilgili ölçeğin psikometrik

özelliklerinin Rasch Ölçme Teorisi çerçevesinde belirlenmesi amaçlanmıştır. Bu amaç doğrultusunda araştırmada cevap aranacak sorular aşağıdaki gibidir.

1. Ölçekte bulunan maddelerin cevaplanması için benimsenen kategori yapısı uygun mudur?
2. Maddelere ve kişilere yönelik güvenilirlik ve ayırt edicilik indeksleri nedir?
3. Maddelerin uyum indeksleri nedir?
4. Ölçek ölçmeyi amaçladığı bireyleri iyi hedeflemekte midir?
5. Maddelerin dağılımı uygun mudur?
6. Ölçek maddeleri yanlılık göstermekte midir?
7. Ölçek tek boyutluluk kriterini sağlamakta mıdır?

1.1. Rasch Ölçme Teorisi

Rasch Ölçme Teorisi (RÖT), temel ölçme işleminin psikolojik değişkenler üzerinde nasıl gerçekleşmesi gerektiğini izah eden bir teorik matematiksel tasvirdir (Bond, Yan ve Heene, 2021). Teori, elde edilen verileri açıklamak için model bulmak yerine ne türden verinin bilimsel ölçme işleminin katı yönergelerine riayet ettiğini belirlemeyi amaçlamaktadır. Bu teoriyi benimseyen araştırmacılar, sosyal bilimlerde ölçmeye ilişkin genel pratiğin ham verilerin tasviri ve ham verilerden yola çıkarak maddeler arasındaki ilişkilerin incelenmesi şeklinde gerçekleştiğini, lakin ham verilerin birer sayım olarak görülmesi gerektiğinden bir ölçüm olarak değerlendirilemeyeceğini savunmaktadırlar (Bond vd., 2021; Boone, Staver ve Yale, 2014). Dolayısıyla, bir testte yer alan maddelerden elde edilen puanların çeşitli aritmetik işlemlere tabi tutularak grupların karşılaştırılması yerleşik bir uygulama olmasına rağmen, bu şekilde elde edilen sonuçların büyük ölçüde yanıltıcı olabileceği görüşündedirler (Wright ve Stone, 1979). Smith, Conrad, Chang ve Piazza (2002), RÖT'ün Klasik Test Teorisine (KTT) nazaran sahip olduğu avantajlardan bir kısmını aşağıdaki gibi belirtmektedirler.

- Verinin modele uyum sergilemesi durumunda ölçülen niteliğin miktarı eşit aralıklı ölçme birimi (logit) ile ifade edilmektedir.
- Veri setindeki kayıpların üstesinden gelme olanağı sunar.
- Hem madde hem de kişi ölçüleri için geçerlik ve güvenilirlik değerlendirmesi sunar.
- Maddelerin örnekleme dağılımından bağımsız olarak kişi yeteneklerine ilişkin tahminde bulunmayı olanaklı kılar.
- Madde ve kişi ölçüleri ortak doğrusal bir ölçek üzerinde ifade edilebilmektedir.
- Grup istatistiklerinden ziyade tekil olarak maddeleri ve kişileri odağına almaktadır.

Sosyal bilimlerde alanında derecelendirme ölçekleri ölçme işlemi için yaygın olarak kullanılan araçlardan biridir. Lakin, KTT bağlamında derecelendirme ölçeklerinden bireylerin elde ettikleri ham puanların doğrudan istatistiksel işlemlere tabi tutulması birçok sınırlılığı beraberinde getirmektedir (Boone vd., 2014). Bunlardan en önemlisi, bu ölçeklerden elde edilen puanların eşit aralıklı olduğunun sorgusuz biçimde kabul edilmesidir (Oon ve Subramaniam, 2011). Genel olarak derecelendirme ölçeklerinde maddeleri cevaplama bireylere seçim yapabilecekleri 2'den fazla kategori sunulmakta ve bu kategoriler ardışık tam sayılar ile puanlanmaktadır. Bu muamele, ölçülen niteliğin miktarına yönelik tek bir madde özelinde ardışık kategorilerin eşit farklılığına, farklı maddelerde ise aynı kategorinin eşit düzeye

işaret ettiği kabulünü beraberinde getirmektedir. RÖT yaklaşımını benimseyen araştırmacılar bu kabulün temelden yoksun olduğunu savunmaktadırlar (Boone ve Staver, 2020). Bu gerekçeden ötürü, derecelendirme ölçeklerinden elde edilen ham verilerin eşit aralıklı değil de sıralama tipinden değerlendirilmesi gerektiğini belirtilmektedirler (Boone vd., 2014). Bunun sonucu olarak; örneğin, 15-75 arasında bir puan elde edilebilecek Likert tipi bir ölçekte, ölçülen niteliğin miktarına yönelik 30 ile 40 puan arasındaki farkın 55 ile 65 puan arasındaki farka eşit olmadığı fikrindedirler.

RÖT bağlamında derecelendirme ölçeklerinin incelenmesi Derecelendirme Ölçeği Modeli (DÖM) ile gerçekleştirilmektedir (Andrich, 1978). Model'in temelinde analizi gerçekleştirilecek derecelendirme ölçeğinin tek bir niteliği ölçtüğü varsayımı bulunmaktadır. Bu nitelik ise ölçeğin Rasch boyutu olarak isimlendirilmektedir. Modelin ölçekte yer alan maddelerin güçlükleri, ölçekte yer alan farklı kategorilerin seçilme olasılıkları ve kişilerin ilgili niteliğe yönelik ölçüleri arasındaki ilişkiyi tesis etmek için kullandığı matematiksel denklem aşağıdaki gibidir (Linacre, 2022). Bu denklem temelinde maddelerin ve kişilerin ölçüleri *logit* terimi ile ifade edilen ölçme birimi ile nicelleştirilmektedir.

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

Yukarıdaki denklemde yer alan F_j değişkeni ardışık kategorilerin seçilme olasılığının eşit olduğu eşik değerleri, B_n değişkeni n 'inci kişinin niteliğe ilişkin ölçüsünü, D_i değişkeni ölçekteki i 'inci maddenin ölçüsünü, P_{nij} ve P_{nij-1} değişkenleri ise bireyin peşi sıra olan kategorileri tercih etme olasılıklarını ifade etmektedir.

Bu model her bir maddenin onaylanma güçlüğünü, testi cevaplayan bireylerin hangi kategoriye tercih ettiğini tahlil ederek belirlemektedir (Bond vd., 2021). Bununla birlikte model, katılımcıların yanıt kategorilerini tercih etme desenini dikkate alarak tüm maddeler için ortak olacak şekilde bir derecelendirme yapısı ortaya çıkarmaktadır (Bond vd., 2021). Bu yapı, ardışık kategoriler arasındaki farkın *logit* birimi ile ifade edildiği eşik değerlerden (τ) oluşmaktadır. Bu eşik değerlerin sayısı kategori sayısının 1 eksiği olup, kategoriler arasında ilerlemenin *logit* birimi cinsinden ne kadarlık artış gerektirdiğini söylemektedir. Kategori eşik değerlerinin işlevselliği Şekil 1'de resmedilen hayali senaryo ile şu şekilde açıklanabilir. Farz edelim ki *Katılmıyorum/Kararsızım/Katılıyorum* şeklinde 3 kategorili bir ölçekte yer alan 7. maddenin ölçüsü 1 *logit*, *Katılmıyorum/Kararsızım* eşik -1.7 ve *Kararsızım/Katılıyorum* eşik 0.8 *logit* olsun. Bununla birlikte Naz, Can, Ayşe ve Ali'nin ölçüleri sırayla -0.7, 1, 1.3 ve 2 *logit* olsun.

cevaplamaya gönüllü olan öğrencilerden Mayıs-Haziran 2022 tarihleri arasında elde edilmiştir. Araştırmaya dahil olan sınıflarda ilk olarak sınıf öğretmenleri testin amacı ve içeriği hakkında bilgilendirilmiştir. Öğretmenlerden öğrencilerine teste verecekleri cevapların gizli kalacağı ve not açısından bir kaygı duymamaları yönünde bilgilendirme yapılmaları istenmiştir.

Tablo 1: Ölçekte Yer Alan Maddeler ve Kodları

Kod	Madde	Kod	Madde
M1-om	Matematiği gerçekten seviyorum	M2	Bazı matematik konularını öğrenmekte zorlanıyorum
M3	Matematiğin anlaşılması çoğu zaman zordur.	M4-om	Matematik eğlenceli bir derstir.
M5-om	Öğretmenim bu ders matematik işleyeceğiz dediğinde mutlu olurum.	M6	Matematik genelde sıkıcı bir derstir.
M7-om	Matematik benim için faydalıdır	M8-om	Keşke okulda daha fazla matematik dersi yapsak!
M9	Matematik derslerinde yardım istemekten çekinirim.	M10	Matematik dersleri beni rahatsız eder.
M11	Matematik derslerinde öğretmenim bana soru soracak diye endişelenirim.	M12	Matematik derslerinde soru sormaktan çekinirim.
M13	Matematik derslerinde tahtaya kalkmaktan korkarım.		

2.2. Araştırmanın Çalışma Grubu

Araştırmada kullanılan veri toplama aracı Trabzon ilinde bulunan okullarda öğrenim görmekte olan 1110 (581 kız, 529 erkek) 3. sınıf, 946 (518 kız, 428 erkek) 4. sınıf olmak üzere toplamda 2056 öğrenciye uygulanmıştır. Veriler toplandıktan sonra yapılan ilk incelemede 195 öğrencinin testten minimum puan (13), 5 öğrencinin ise maksimum puan (39) elde ettiği belirlenmiştir. Bu durum 195 öğrencinin kaygı seviyesinin testin belirleyebileceği minimum kaygı seviyesine eşit ya da daha az olduğunu, 5 öğrencinin kaygı seviyesinin ise testin belirleyebileceği maksimum kaygı seviyesine eşit ya da daha fazla olduğunu söylemektedir. DÖM her ne kadar bir testten maksimum ya da minimum puanı elde eden bireyler için bir ölçüm kestirimi sunsa da bu kişilerin ilgili niteliğe ilişkin ölçülerinin ne olduğunun tam olarak belirlenmesi mümkün değildir. Bu sebepten ötürü, RÖT çerçevesinde 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 olmasından ötürü testin ne kesinlikte işlev gösterdiğini anlamak için faydalı bilgi sunmamasıdır (Boone vd., 2014). Bu gerekçeden hareketle bu kişiler veri analizine dahil edilmemiştir. Sonuç olarak araştırmanın bulguları 993 (535 kız, 458 erkek) 3. sınıf, 863 (476 kız, 387 erkek) 4. sınıf olmak üzere toplamda 1856 öğrenciden elde edilmiştir.

2.3. Veri Analizi

Öğrencilerin testte yer alan maddelere verdikleri yanıtlar daha önce açıklandığı biçimde puanlandırıldıktan sonra araştırma sorularına cevap vermek için gereken analizler Winsteps® (sürüm 3.91) yazılımı ile gerçekleştirilmiştir.

3. BULGULAR

Bu bölümde, araştırma çerçevesinde elde edilen bulgular cevap aranan problemlerin belirlediği alt başlıklar hâlinde sunulmuştur.

3.1. Kategori Yapısının İşlevselliğine Yönelik Bulgular

Derecelendirme türünden ölçekler maddelere yönelik ikili yanıt seçeneklerinin sunabileceğinden daha detaylı bilgiye ulaşmak için kullanılmaktadır. Bununla birlikte, maddeleri yanıtlamada daha fazla kategorinin katılımcılara sunulması, ölçülmesi hedeflenen değişkene yönelik daha geçerli ve hassas verinin elde edildiğini garanti etmemektedir. Bu hususta Andrich (1996), derecelendirme ölçeklerini cevaplandırırken katılımcıların yanıt kategorilerini testi hazırlayan araştırmacıların kurguladıklarından farklı biçimde kullanabildiklerini belirtmektedir. Dolayısıyla, yanıt kategorilerinin sayısının artırılarak arzu edilen amaca ulaşıp ulaşılmadığı, kategorilerin maddeler ile örtük değişken arasında kurgulanan ilişkiyi açığa çıkarmada işlevsel olup olmadığının analiz edilmesi gerekmektedir.

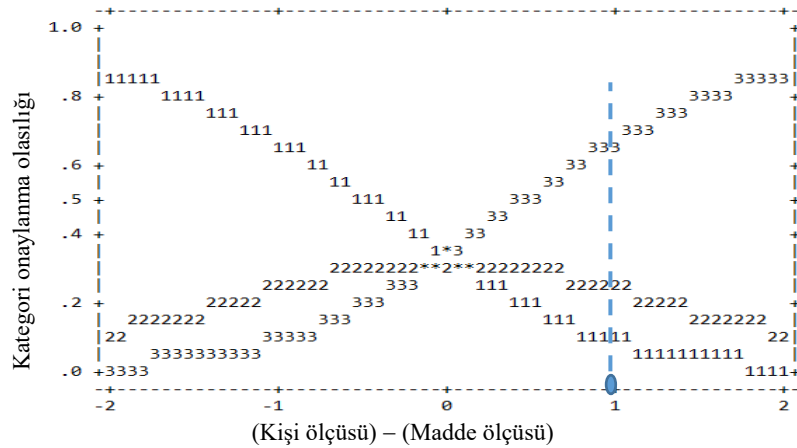
Linacre'nin (2002) bir ölçekte kullanılan kategorilerin yapısının işlevselliğini sınamak için önerdiği kıstaslar şu şekildedir; a) her bir kategoride en az 10 yanıt bulunmalıdır, b) kategorilerdeki ölçülerin ortalamaları monoton artan olmalıdır, c) kategorilerin dış uyum indeksleri 2'den küçük olmalıdır, d) Ardışık kategori eşikleri monoton artan olmalı ve aralarındaki fark 1.4 logit değerinden büyük fakat 5 logit değerinden küçük olmalıdır, e) Kategori olasılık eğrilerinin her biri zirve yapmalıdır. Linacre'nin (2002) ilk 4 önerisine ilişkin elde edilen bulgular Tablo 2'de sunulmuştur.

Tablo 2: Kriterlere İlişkin Elde Edilen Bulgular

Kategoriler	Frekans (%)	Ortalama ölçü	Dış uyum (MnSq)	Kategori eşik (Andrich)
1 Puan	14825 (61)	-1.8	1.14	
2 Puan	4513 (19)	-0.42	0.71	0.04
3 Puan	4790 (20)	0.52	1.16	-0.04

Tablo 2'den görüldüğü üzere her bir kategoride 10'dan fazla yanıt bulunmakta ve kategorilerdeki ölçülerin ortalamaları monoton artandır. Ayrıca, dış uyum istatistikleri 2'den küçüktür. Kategori eşik değerleri ise monoton artan değildir ve aralarındaki fark önerilen aralıkta değildir.

Kategorilerin işlevselliğine ilişkin son kriteri değerlendirmek için oluşturulan kategori olasılık eğrileri Şekil 2'de görülmektedir.



Şekil 2: Kategori Olasılık Eğrileri

Şekil 2’de resmedilen yapıda yatay eksen bir kişinin matematik kaygısı ölçüsünden bir maddenin ölçüsü çıkartıldığında elde edilen farkı temsil etmektedir. Düşey eksen ise bir kategorinin onaylanma olasılığını belirtmektedir. Örneğin, bir maddenin ölçüsü 1 logit ve testi cevaplayan bir öğrencinin kaygı ölçüsü 2 logit olsun. Bu durumda ikisi arasındaki fark 1 logit olup, şekil 2’de çizilen doğru parçasının rakamlar ile ifade edilmiş eğrileri kestiği noktaların ordinat değerleri bu kişinin bu maddede kategorileri onaylama olasılıklarını belirtmektedir. Dolayısıyla, bu kişi bu maddede 3 puan elde edeceği kategoriyi onaylama olasılığı yaklaşık %65, 2 puanlık kategoriyi onaylama olasılığı yaklaşık %30, 1 puanlık kategoriyi onaylama olasılığı ise yaklaşık %15’tir. Şekil 2’den görüldüğü üzere, kişi ve madde kombinasyonlarının tümü için 2 puanlık kategori hiçbir zaman için seçilmesi en olası kategori değildir. Linacre’nin ifadesiyle, 2 puanlık kategoriyi temsil eden eğri zirveye sahip değildir.

3.2. Güvenirlik ve Ayırt Edicilik İndekslerine İlişkin Bulgular

RÖT bağlamında kişilere yönelik elde edilen güvenilirlik indeksi Cronbach alfa katsayısına benzer biçimde yorumlanmaktadır (Linacre, 2022). Elde edilen değer 1’e ne kadar yakınsa yapılan ölçüm o derece yüksek iç tutarlılığa sahiptir. Bir başka ifadeyle kişi güvenilirlik indeksinin yüksek olması, bazı kişilerin yüksek bazı kişilerin düşük puan alacak şekilde bir test oluşturulduğunu ve kişilere ilişkin hesaplanan ölçülerin tutarlı olduğu anlamına gelir (Bond vd., 2021). Kişi güvenilirlik indeksi yanı sıra Rasch analizi maddelere ilişkin de güvenilirlik indeksi sunmaktadır. Maddelere ilişkin güvenilirlik değeri, ölçülmesi hedeflenen değişkenin farklı seviyeleri boyunca maddelerin kendi içerisinde bir hiyerarşi tanımlama gücüne işaret etmektedir (Linacre, 2022). Bir analogi ile örneklendirilecek olursa, elimizde yalnızca 15C° ile 25C° arasındaki sıcaklık ölçülerini belirleyebilen bir termometre olduğunu düşünelim. Eğer sıcaklıklarını ölçmek istediğimiz nesnelerin sıcaklıkları 1C° ile 50C° arasında değişiyorsa termometre bu nesnelerin bir bölümünün sıcaklık değerlerini hassas biçimde ölçemeyecektir. Dolayısıyla maddelerin güvenilirliği düşükse, test ölçülecek değişkenin dar bir aralığına hitap etmektedir. Gerçekleştirilen ölçüm sonucunda ham puanların güvenilirlik indeksi (Cronbach α) .86, kişilerin logit ölçülerinin güvenilirlik indeksi .72, maddelerin logit ölçülerinin güvenilirlik indeksi ise .99 olarak hesaplanmıştır. Hesaplanan kişi ve madde güvenilirlik indeksleri literatürde önerilen sınır olan .7’den büyüktür (Boone vd., 2014).

Ayırt edicilik indeksi (G) gerçek puanların standart sapmasının hataların standart sapmasına oranıdır (Linacre, 2022). Bir başka ifadeyle örneklemin, ölçme kesinliği temelinde dağılımını belirtmektedir (Wright ve Stone, 2004). Kişilere yönelik ayırt edicilik indeksi ölçüme tabi olan bireylerin ölçülen nitelik açısından sınıflandırılmasına ilişkin bilgi sunmaktadır (Bond vd., 2021). Düşük ayırt edicilik indeksi, ölçmede kullanılan aracın kişileri ölçülen niteliğe sahiplik açısından ayırtırmada yeterli olmadığına işaret etmektedir. Kişi ayırt edicilik indeksi 3'ten büyük ise mükemmel, 2-3 arasında ise iyi, 1.5'ten büyük ise kabul edilebilir olarak değerlendirilmektedir (Duncan, Bode, Lai ve Perera, 2003). Madde ayırt edicilik indeksinin ise birey bazında analiz gerçekleştirebilmek için en az 1.5, grup bazında analiz için ise 2.5'ten büyük olması önerilmektedir (Boone vd., 2014). Bu çalışmada kişi ayırt edicilik indeksi 1.6, madde ayırt edicilik indeksi ise 20.96 olarak belirlenmiştir. Yapılan bir ölçüm sonucunda elde edilen kişi ayırt edicilik indeksi ile bireylerin ilgili niteliğe sahiplik açısından kaç farklı tabakaya ayrılacağı aşağıdaki formülle saptanmaktadır (Linacre, 2022). Bu çalışmada elde edilen kişi ayırt edicilik değeri yerine yazıldığında tabaka değeri yaklaşık olarak 2.5 elde edilmiştir.

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

3.3. Madde Uyum İndekslerine İlişkin Bulgular

Gerçekleştirilen bir Rasch analizi sonucunda eş aralıklı ölçülerin elde edilebilmesi verilerin modele uyum sergilemesi durumunda mümkün olmaktadır (Bond vd., 2021). Rasch analizi hem kişilere hem de maddelere yönelik uyum istatistikleri sunmaktadır. Bir ölçekte Rasch modeline uyum sergilemeyen maddelerin hedeflenen örtük değişkenin ölçümüne hizmet etmediği ifade edilmektedir (Oon ve Subramaniam, 2011). Bununla birlikte Wright ve Linacre (1994) kişilerin belirli bir niteliğe ilişkin sahiplik düzeylerinin farklı zamanlarda değişkenlik gösterebileceğini, lakin bir testte yer alan maddelerin güçlüklerinin farklı zamanlarda değişmesinin ölçme açısından istenmeyen bir durum olduğunu gerekçe göstererek, kişilere nazaran madde uyum indekslerinin daha ön planda olması gerektiğini vurgulamaktadır.

Tablo 3: Madde Uyum İstatistikleri ve Nokta-Ölçü Korelasyon Değerleri

Madde	Madde Ölçüsü (SH)	Dış uyum (MnSq)	İç uyum (MnSq)	Nokta-ölçü korelasyonu
M7-om	2.03 (0.08)	1.25	1.37	0.29
M10	1.02 (0.05)	0.63	0.86	0.56
M1-om	0.57 (0.04)	0.84	0.79	0.58
M4-om	0.54 (0.04)	0.76	0.84	0.59
M6	0.51 (0.04)	0.66	0.85	0.62
M13	0.19 (0.04)	0.80	0.97	0.62
M5-om	0.01 (0.04)	0.83	0.85	0.60
M12	-0.15 (0.04)	1.03	1.16	0.57
M11	-0.16 (0.04)	0.88	1	0.62

M8-om	-0.65 (0.03)	1.22	1.06	0.49
M9	-0.66 (0.03)	1.42	1.30	0.50
M3	-1.33 (0.03)	1.17	1.07	0.56
M2	-1.92 (0.04)	1.48	1.13	0.49

SH: Standart Hata

RÖT çerçevesinde uyum indeksleri artık kareler ortalaması (Mean-square residuals “MnSq”) üzerinden hesaplanmaktadır (Boone vd., 2014). İç uyum (Infit) ve dış uyum (Outfit) ismiyle iki farklı türden hesaplanan bu değerler birer Ki-kare istatistiği olup veriler ile model arasındaki uyumun ne düzeyde olduğunu ifade etmektedir. Bu istatistiklerin beklenen değeri 1 olup pozitif her gerçek sayı değerini alabilmektedir (Wright ve Linacre, 1994). Hesaplanan değer 1’den büyük ise yanıtların deseninin model açısından daha az öngörülebilir olduğunu, 1’den küçük ise daha çok öngörülebilir olduğunu ifade etmektedir. Örneğin, bir uyum indeksinin değeri 1,48 ise verilerde DÖM’nin beklentisinden %48 daha fazla varyans, 0,68 ise %32 daha az varyans bulunmaktadır. Bir maddenin dış uyum indeksi, ölçüleri o maddenin ölçüsünden uzakta olan bireylerin yanıtlarına, iç uyum indeksi ise ölçüleri maddenin ölçüsüne yakın olan bireylerin yanıtlarına hassastır (Bond ve Fox, 2007). Wright ve Linacre (1994) derecelendirme ölçeği için uyum indekslerine yönelik makul aralığı 0.6-1.4 olarak belirtmekte birlikte 0.5-1.5 aralığının ölçme işlemi için yararlı olduğunu not etmişlerdir.

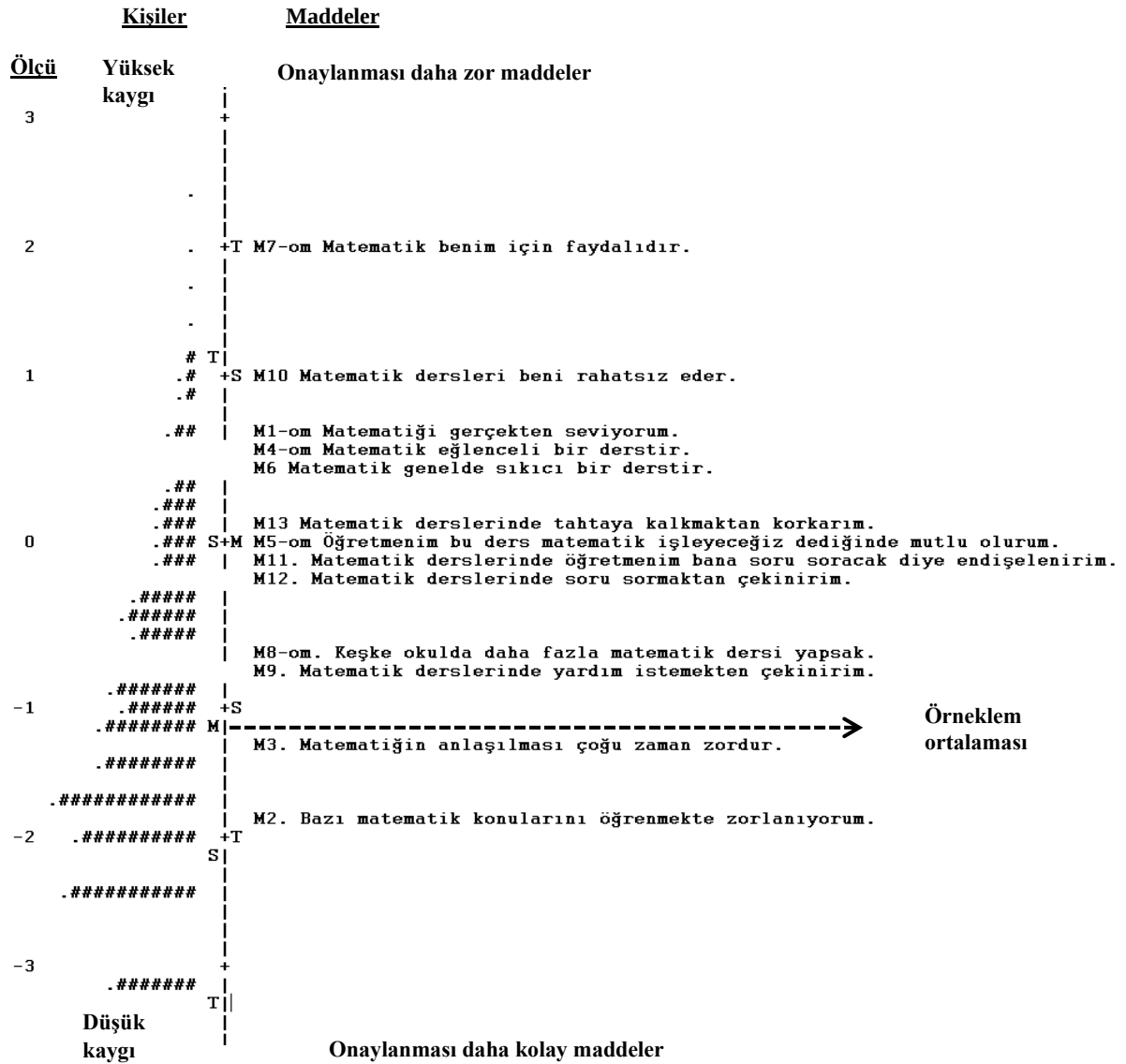
Bir testte yer alan maddelerin, ölçülmesi hedeflenen örtük değişken ile uyumlu olup olmadığını değerlendirmede göz önüne alınması önerilerin bir diğer husus nokta-ölçü korelasyonudur (Boone ve Staver, 2020). Sıfıra yakın ya da negatif değerler ilgili maddenin testin bütünü ile ters düştüğüne işaret etmektedir (Bond vd., 2021). Bu değer, bir maddeye verilen yanıtlar ile yanıtları veren bireylerin ölçüleri arasındaki ilişkidir (Bond vd., 2020). Bu çalışmada testten elde edilen puan -dolayısıyla kişi ölçüsü-artıkça kaygı düzeyinin arttığı kurgulandığından, maddelere verilen yanıtlardan elde edilen puan ile kişi ölçüleri arasındaki korelasyon pozitif olması gerekmektedir. Nokta-ölçü korelasyon değerinin 0.3’ten büyük olması önerilmektedir (Boone, 2020). Gerçekleştirilen analiz sonucunda elde edilen maddelerin uyum indeksleri, nokta-ölçüm korelasyon değerleri ve standart hataları ile birlikte madde ölçüleri Tablo 3’te sunulmuştur. Tablo 3’te maddeler yukarıdan aşağıya ölçüleri (logit) azalacak biçimde sıralanmıştır. Maddenin ölçüsünün artması onaylanmasının zorlaştığına, diğer bir ifadeyle ölçüleri kendisinininkinden daha küçük maddelere kıyasla matematik kaygısı değişkeninin daha üst seviyesine hitap ettiği anlamına gelmektedir.

3.4. Madde Dağılımı ve Örneklemenin Hedeflenmesine İlişkin Bulgular

Bir ölçeğin kapsam geçerliliğini arttıran etmenlerden biri ölçek içerisinde yer alan maddelerin ölçülmesi hedeflenen değişkenin farklı seviyelerine hitap edebilmesidir (Boone vd., 2014). Şayet ölçek içerisindeki maddelerin ölçüleri birbirine çok yakın ise ölçek bir bütün olarak ölçmeyi hedeflediği değişkenin düşük ve yüksek seviyelerini belirlemede hizmet etmeyecektir. RÖT bağlamında bir testte yer alan maddelerin bu yeterliliğini incelemek için kullanılan görsel araç kişi-madde haritasıdır (Wright-map). Bu harita maddelerin ve kişilerin birbirine göre konumlarını düşey bir doğru parçasının iki tarafında görselleştirmektedir. Bu düşey doğru parçası ölçülmesi hedeflenen değişkenin miktarını belirleyen bir cetvel olarak düşünülebilir. Bu şekilde ele alındığında maddeler de cetvelin üzerindeki sayılar olacaktır.

Kişi madde haritası Şekil 3'te sunulmuştur. Şeklin ortasındaki doğru parçasının solunda, örneklemedeki bireylerin matematik kaygı ölçülerine göre dağılımı yer almaktadır. Bu dağılımda “#” sembolü 16 kişiyi, “.” sembolü ise 1-15 arasında bir tam sayı kadar kişi sayısını temsil etmektedir. Doğru parçasının her iki tarafında yer alan M, S ve T harfleri sırasıyla kişilerin ve maddelerin ölçüleri dağılımlarının ortalamasını, 1 standart sapma sınırını ve 2 standart sapma sınırını temsil etmektedir. Kişiler kısmında aşağıdan yukarıya gidildikçe kaygı düzeyi artmaktadır. Maddeler kısmında ise aşağıdan yukarı gidildikçe maddelerin onaylanması zorlaşmaktadır, diğer bir değişle üst kısımlarda yer alan maddelerden 3 puan elde edilmesi aşağıdaki maddelere nazaran daha yüksek kaygı gerektirmektedir.

Maddelerin konumları dikkate alındığında kısa bir aralık içerisinde hapsolmadıkları, dolayısıyla testin bir bütün olarak matematik kaygısı değişkeninin farklı seviyelerine hitap ettiği söylenebilir. Test içerisinde en yüksek kaygı seviyesini M7-om kodlu maddenin, en düşük kaygı seviyesini ise M2 kodlu maddenin ölçtüğü görülmektedir. Dolayısıyla düşük kaygı seviyesine sahip öğrenciler de matematikteki konuları öğrenmede zorlandıklarını belirtmişlerdir. Diğer yandan, yalnızca çok yüksek kaygı seviyesine sahip olan öğrenciler matematiğin faydalı olduğuna katılmamaktadır. Maddeler arasındaki boşluklar dikkate alındığında, testin en zor iki maddesi arasında yaklaşık 1 logit kadar farklılık bulunmaktadır. Bununla birlikte M4-om ile M6 kodlu maddeler hemen hemen aynı ölçüye sahip olup bu maddeler anlamsal olarak çok benzerdir.



Şekil 3: Kişi-Madde Haritası

Kişilerin bir niteliğe ilişkin sahiplik düzeylerini belirlemek için uygulanan bir testin örnekleme iyi hedeflemesi bir gerekliliktir (Wright ve Stone, 1979). Testin hedefleme özelliğinin iyi olması, testin uygulandığı kişiler için ne çok zor ne de çok kolay olmaması anlamına gelmektedir (Boone vd., 2014). Örnekleme iyi hedefleyen bir testin uygulanması sonucunda elde edilen kişi-madde haritasında maddelerin yoğunlaştığı bölge ile kişilerin yoğunlaştığı bölgenin karşılıklı olması beklenir (Bond ve Fox, 2007). Kişi ve maddelerin dağılımı bu açıdan ele alındığında, Şekil 3'te görüleceği üzere kişilerin öbeklendiği bölge maddelere nazaran daha alt seviyede oluşmuştur. Niceliksel olarak hedefleme yeterliliğine ilişkin Duncan vd. (2003), kişilerin ölçülerinin ortalaması maddelerin ortalamasından 0.5 logit kadar farklıysa hafif düzeyde isabetsizliği, 1 olmasının ise önemli bir isabetsizliği gösterdiğini belirtmektedir. Rasch analizi maddelerin ölçüleri ortalamasını 0 logit değerine kalibre etmektedir. Çalışmada kişilerin ölçüleri ortalaması -1.08 logit, standart sapması ise 1.11 logit olarak hesaplanmıştır. Bu durum ölçeğin örnekleme hedeflemeye kayda değer bir isabetsizlik sergilediğini göstermektedir.

3.5. Maddelerin Yanlılığına İlişkin Bulgular

Madde işleyiş farklılaşması analizi (MİFA) bir testin içerisinde yer alan maddelerin örneklemin içerisinde yer alan farklı gruplarda aynı şekilde işleyip işlemediğini tahlil etmektedir. RÖT bağlamında bir maddenin işleyişindeki farklılaşma, ölçüleri aynı olan fakat farklı gruplarda bulunan bireylerin bir maddeye yanıt verme olasılıkları arasında istatistiksel olarak anlamlı bir fark bulunduğunda ortaya çıkmaktadır (Linacre, 2022). Örneğin, matematiğe ilişkin yüksek düzeyde kaygıya karşılık gelen bir maddede, kaygı seviyesi yüksek olan kadınlar çoğunlukla “katılıyorum” kategorisini onaylamış fakat kaygı seviyesi yüksek olan erkekler çoğunlukla “katılmıyorum” kategorisini onaylamış ise madde cinsiyet değişkeni açısından işleyiş farklılaşması sergiliyor anlamına gelir. Diğer bir değişle, her iki grupta oluşturulan kişi-madde haritalarında madde hiyerarşisinin anlamlı derecede farklılaşmasıdır (Boone vd., 2014).

Matematik kaygısına ilişkin gerçekleştirilen araştırmalarda öğrencilerin cinsiyet ve sınıf düzeyi açısından farklılık gösterip göstermediği sıklıkla inceleme konusu olduğu görülmektedir. Dolayısıyla, bu çalışmada incelen testin bu değişkenler açısından yanlılık sergileyip sergilemediği analiz edilmiştir. Boone vd. (2014), maddelerin bir sınıflama değişkeni açısından işleyiş farklılaşması gösterip göstermediğine karar kılmak için maddelerin ölçüleri arasındaki farkın mutlak değerinin 0.64 logit değerinden büyük ve olasılık değerinin de .05’ten küçük olmasını kriter olarak sunmaktadır.

Tablo 4: Cinsiyet Değişkenine Göre Gerçekleştirilen MİFA Bulguları

Madde	Kadın		Erkek		MİF fark (Kadın-Erkek)	t	p
	MİF ölçüsü	SH	MİF ölçüsü	SH			
M7-om	2.21	0.11	1.8	0.11	0.41	2.64	.01
M10	1.05	0.07	.99	0.08	0.06	.56	.58
M1-om	0.61	0.06	.51	0.07	0.10	1.05	.29
M4-om	0.59	0.06	0.47	0.07	0.12	1.38	.17
M6	0.57	0.06	0.42	0.06	0.15	1.74	.08
M13	0.11	0.05	0.30	0.06	-0.19	-2.33	.02
M5-om	0.06	0.05	-0.06	0.06	0.12	1.53	.13
M12	-0.18	0.05	-0.12	0.06	-0.06	-0.86	.39
M11	-0.23	0.05	-0.06	0.06	-0.17	-2.28	.02
M8-om	-0.60	0.05	-0.72	0.05	0.12	1.71	.09
M9	-0.59	0.05	-0.75	0.05	0.16	2.25	.02
M3	-1.36	0.05	-1.30	0.05	-0.06	-0.80	.43
M2	-2.07	0.05	-1.75	0.05	-0.32	-4.52	.00

Cinsiyet değişkeni odağa alınarak gerçekleştirilen MİFA sonucunda elde edilen bulgular Tablo 4’te sunulmuştur. Gerçekleştirilen analizde her bir madde için “Madde ölçüleri iki grupta eşittir” hipotezi test edilmiştir.

Tablo 4’ten görüldüğü üzere, iki gruptaki ölçüleri arasındaki farkın en büyük olduğu iki madde M7-om ve M2 kodlu maddelerdir. Bununla birlikte M7-om kodlu madde her iki grup için de onaylanması

en zor, M2 kodlu madde ise her iki grup için de onaylanması en kolay madde olarak konumlarını korumaktadır. Boone vd.'nin önerdiği kriterler temelinde hiçbir madde cinsiyet değişkeni açısından işleyiş farklılaşmasına sahip değildir.

Sınıf değişkeni açısından gerçekleştirilen MİFA sonucunda elde edilen bulgular Tablo 5'te sunulmuştur. Analizde her bir madde için “Madde ölçüleri iki grupta eşittir” hipotezi test edilmiştir. Tablo 5'te sunulan değerler sınıf seviyesi açısından maddelerin işleyişinin farklılaşmadığına işaret etmektedir.

Tablo 5: Sınıf Değişkenine Göre Gerçekleştirilen MİFA Bulguları

Madde	3. sınıf		4. sınıf		MİF fark (3-4)	t	p
	MİF ölçüsü	SH	MİF ölçüsü	SH			
M7-om	1.99	0.11	2.06	0.11	-0.07	-0.46	.64
M10	1.02	0.07	1.02	0.07	0.00	0.00	1
M1-om	0.65	0.06	0.48	0.06	0.17	1.98	.05
M4-om	0.60	0.06	0.48	0.06	0.12	1.46	.14
M6	0.54	0.06	0.47	0.06	0.08	0.87	.38
M13	0.22	0.06	0.16	0.06	0.06	0.80	.43
M5-om	0.09	0.05	-0.08	0.06	0.17	2.18	.03
M12	-0.19	0.05	-0.12	0.05	-0.07	-0.93	.35
M11	-0.16	0.05	-0.16	0.05	0	0	1
M8-om	-0.63	0.05	-0.68	0.05	0.05	0.65	.52
M9	-0.69	0.05	-0.63	0.05	-0.06	-0.91	.36
M3	-1.37	0.05	-1.28	0.05	-0.09	-1.41	.16
M2	-2.03	0.05	-1.78	0.05	-0.26	-3.57	.001

3.6. Ölçeğin Boyut Yapısına İlişkin Bulgular

RÖT'nin varsayımlarından biri tek boyutluluk (unidimensionality) ilkesidir (Linacre, 2022). Bu ilke, bir testte kişilerin gösterdikleri performansın yalnızca bir örtük değişken tarafından açıklanabilmesi olarak tanımlanmaktadır (Brentari ve Silvia, 2007). Bu örtük değişken testin Rasch boyutu olarak isimlendirilmektedir (Boone ve Staver, 2020). RÖT bağlamında bir testin tek boyutluluk kriterini sağlayıp sağlamadığının tespiti PCAR (Principal Component Analysis of Residuals) analizi ile gerçekleştirilmektedir (Boone ve Staver, 2020). Bu analizde artıklar (residuals), gözlenen puanların beklenen değerlerinden farkını ifade etmektedir (Wright ve Masters, 1990). PCAR verideki toplam varyanstan Rasch boyutundan kaynaklı varyansı çıkardıktan sonra geriye kalan varyansın rasgele hatalardan ileri gelip gelmediğini tahlil eder. Bu tahlil için gerçekleştirilen faktör analizinin amacı, artıkların varyansını en çok açıklayan faktörün, böyle bir faktörün var olduğu varsayımı altında, ortaya çıkarılmasıdır (Linacre, 1998). Eğer elde edilen faktörün kaynağı verideki rasgele gürültü (random noise) olduğu tespit edilirse artıklar içerisinden belirgin bir faktörün mevcut olmadığı çıkarımı yapılır (Linacre, 1998).

PCAR işlemi sonrasında artıkların korelasyon matrisinden elde edilen birinci temel bileşenin özdeğeri 2'den küçük ise Rasch boyutu dışında kalan varyansın rasgele hatalardan ileri geldiği sonucuna

varılır (Boone ve Staver, 2020). Bununla birlikte Rasch boyutundan kaynaklı varyans toplam varyansın %50'den fazlasını açıklıyorsa bu durumun da tek boyutluluk ilkesinin sağlandığının bir başka göstergesidir (Boone ve Staver, 2020).

Tablo 6: Madde Kümeleri ve Kümeler Arası Korelasyon Değerleri

Madde Kümeleri			Kümeler arası korelasyon değerleri					
1	2	3	1-2		1-3		2-3	
			Pearson	Azaltılmamış	Pearson	Azaltılmamış	Pearson	Azaltılmamış
M1-om, M5-om, M4-om, M8-om, M7-om, M6	M10, M3, M2	M12, M9, M13, M11	0.39	0.99	0.31	0.93	0.41	0.99

Gerçekleştirilen bir PCAR neticesinde maddeler, kendi aralarında sergiledikleri korelasyon büyüklüğüne göre 3 farklı kümeye ayrılmaktadır. Kümelerin birbiri arasındaki korelasyon katsayısının 1'e yakın bir değer çıkması farklı kümelerdeki maddelerin aynı yapıyı ölçtüğüne işaret etmektedir (Linacre, 2022). Çalışmada gerçekleştirilen PCAR neticesinde ortaya çıkan madde kümeleri Tablo 6'da sunulmuştur. Tablo 6 içerisinde yer alan azaltılmamış korelasyon (disattenuated correlation) katsayısı, gerçekleştirilen ölçme işleminin hatalardan arınık olması durumunda ortaya çıkacak korelasyon katsayısını göstermektedir. Tablo 6'da 1. ve 3. kümelerde yer alan maddelere dikkat edildiğinde olumlu ve olumsuz maddeler ayrı ayrı kümelenmişlerdir.

Bu çalışmada gerçekleştirilen PCAR sonucundan testin Rasch boyutunun açıkladığı varyansın toplam varyanstaki oranı %47 ve birinci bileşenin özdeğeri 2.34 olarak belirlenmiştir. Elde edilen bu değerler önerilen sınırlardan biraz farklılık göstermektedir. Linacre (2022), böyle bir durumla karşılaşıldığında 1. ve 3. kümeler arasındaki azaltılmamış korelasyon katsayısını dikkate almayı ve bu değer 0.7'den büyük ise iki kümedeki maddelerin aynı yapının ölçülmesine hizmet ettiğine karar kılınabileceğini belirtmektedir. Tablo 6'dan görüldüğü üzere 1. ve 3. kümeler arasındaki azaltılmamış korelasyon katsayısı 0.93 olup önerilen eşiğin üzerindedir.

4. TARTIŞMA ve SONUÇ

Çalışmada incelenen hususlardan ilki testin maddelerine cevap vermede benimsenen 3'lü kategori yapısının işlevsel olup olmadığını belirlemektir. Elde edilen bulgular, benimsenen kategori yapısının literatürde kategorilerin seçilme frekansı, kategorilerin ortalama ölçüleri ve kategori uyum indekslerine yönelik önerilen kriterleri sağladığını fakat kategori eşik değerlerine ve seçilme olasılıklarına ilişkin kriterleri sağlamadığını göstermiştir. Bunun sonucunda *kararsızım* kategorisi hiçbir madde-kişi kombinasyonu için seçilme olasılığı en yüksek kategori olmamıştır. Bununla birlikte 2 puanlık kategori eşiğinin değeri 1 puanlık kategori eşiğinin değerinden küçüktür. Bu durum, bir maddenin işaret ettiği kaygı seviyesinden daha düşük seviyede kaygıya sahip olan bir bireyin ilgili maddede 3 puanlık kategoriye onaylama olasılığının 2 puanlık kategoriye onaylama olasılığından büyük olduğunu, daha yüksek seviyede kaygıya sahip bir bireyin ise ilgili madde için 1 puanlık kategoriye onaylama olasılığının 2 puanlık kategoriye onaylama olasılığından büyük olduğunu söylemektedir. Bu ise testten elde edilen puan ile kaygı seviyesi arasında kurulan ilişki ile zıtlık oluşturmaktadır. Boone vd. (2004) testlerde

kararsızım kategorisinin bir örnek olduğu tarafsız bir kategorinin kullanılmamasının ölçme açısından daha faydalı ve üretken sonuçlar ortaya çıkardığını belirtmektedir. Bu çalışmada elde edilen bulgular bu sonucu destekler niteliktedir. Dolayısıyla, ileride burada incelenen kaygı ölçeğinden faydalanılacak çalışmalarda *kararsızım* kategorisinin kaldırılarak *Kesinlikle katılıyorum – Katılıyorum – Katılmıyorum – Kesinlikle katılmıyorum* şeklinde 4'lü bir kategori yapısının kullanılması daha işlevsel bir ölçmenin gerçekleşmesine imkân tanıyabilir.

Araştırmada ele alınan bir diğer husus yapılan ölçümün güvenilirlik ve ayırt edicilik özelliklerinin incelenmesiydi. Bireylerin testten elde ettikleri ham puanlar üzerinde hesaplanan güvenilirlik katsayısı ölçümün yüksek düzeyde iç tutarlılık sergilediğini ortaya koymuştur. Rasch analizi sonucunda elde edilen kişi ve madde ölçüleri üzerinde hesaplanan güvenilirlik indeksleri ise literatürde belirtilen eşik değerden yüksektir. Bu durum, testin farklı bir örneklem üzerinde uygulanması ile elde edilecek madde ölçülerinin çok büyük ihtimalle bu çalışmada hesaplanan değerlere yakın olacağı, bir diğer anlamda test maddelerinin kişi-madde haritasındaki konumlarının aynı kalacağına işaret etmektedir (Bond vd., 2021). Yapılan analiz sonucunda maddelere ilişkin hesaplanan ayırt edicilik indeksi literatürde önerilen eşik değerden yüksek çıkmıştır. Buna bağlı olarak maddelere yönelik tabaka değeri de oldukça yüksektir. Bu durum, maddelerin bir bütün olarak testin ölçmeyi hedeflediği matematik kaygısı değişkeninin geniş bir aralığını ölçme yeterliliğine sahip olduğuna işaret etmektedir. Diğer taraftan, kişilere yönelik ayırt edicilik indeksi ise madde ayırt edicilik indeksine kıyasen daha düşüktür. Buna bağlı olarak kişilere yönelik hesaplanan tabaka değeri yaklaşık olarak 2.5 elde edilmiştir. Dolayısıyla test, kişileri matematik kaygısına sahiplik açısından en azından düşük-orta-yüksek şeklinde 3 tabakaya ayıramamaktadır. Ortaya çıkan bu sonuç testin hedefleme yeterliliği ile doğrudan ilişkilidir. Elde edilen bulgular, testin uygulandığı örneklemi hedefleme hususunda kayda değer bir isabetsizlik sergilediğini ortaya koymuştur. Kişilerin ve maddelerin dağılımı dikkate alındığında test bu örneklem için onaylanması zordur. Buradaki zorluktan kasıt, testte yer alan maddelerin birçoğunun işaret ettiği matematik kaygı seviyelerinin örneklemde yer alan bireylerin çoğunun kaygı seviyesinden yüksek olmasıdır. Şekil 3'te sunulan kişi-madde haritasından görüleceği üzere, ölçüsü örneklem ortalamasından küçük olan yalnızca iki madde (M3 ve M2) bulunmaktadır. Yine haritadan görüleceği üzere, örneklemdeki kayda değer büyüklükteki katılımcının ölçüsü -0.8 logit değerinden küçük iken ölçüsü bu değerden küçük yalnızca bu iki madde bulunmaktadır. Dolayısıyla, testin maddeleri örneklem bu kısmında yer alan bireyleri sahip oldukları matematik kaygısı açısından ayırt etmede yeterince hizmet edememektedir. Bu açıdan ele alındığında, ölçeğin matematik kaygısı değişkeninin daha düşük seviyelerine hitap edebilecek maddeler ile zenginleştirilmesinin gerekli olduğu değerlendirilmektedir.

Şekil 3'te yer alan kişi-madde haritasından görüldüğü üzere testin onaylanması en zor maddesi M7-om kodlu maddedir. Bu madde ile onaylanma zorluğu açısından ikinci sırada yer alan M10 kodlu madde arasında yaklaşık 1 logit büyüklüğünde farklılık bulunmaktadır. Kişi-madde haritasında bireylerin dağılımı dikkate alındığında M7-om kodlu maddenin seviyesinde çok az kişi bulunmaktadır. Daha açık olarak, bu maddenin ölçüsü 2.03 logit olmakla birlikte 1856 katılımcı arasında ölçüsü 2 logitten büyük veya eşit 13 katılımcı olup sıralamada 14. kişinin ölçüsü 1.68 logittir. Dolayısıyla M7-om kodlu madde bireyleri kaygı düzeyi açısından ayırt etmede işlevsel görünmemektedir. Ayrıca, Tablo 3'te sunulduğu üzere bu maddenin nokta-ölçü korelasyon değeri tüm maddeler içerisinde en küçük olup Boone'un (2020) önerdiği eşiğin altındadır. Bu ise maddenin testin bütünüyle birlikte hareket etmediğini, kaygı değişkenine ait bir madde olmayabileceğine işaret etmektedir. Tüm bunlardan hareketle ileriki araştırmalarda bu maddenin testten atılması değerlendirilebilir. Maddelerin dağılımı açısından göze çarpan bir diğer husus M4-om kodlu madde ile M6 kodlu maddeler için elde edilen ölçülerdir. Tablo

3'ten görüleceği üzere iki maddenin ölçüleri çok yakın olmanın yanında iki madde anlamsal açıdan da birbirine çok benzerdir. Dolayısıyla ileriki araştırmalarda bu maddelerden biri testten atılabilir.

RÖT'ün temel varsayımı olan tek boyutluluk, geliştirilen bir testin yalnızca tek bir örtük değişkenin ölçümüne hizmet etmesi gerektiği anlamına gelmektedir. Bu çalışmada incelenen testin tek boyutlu olup olmadığını incelemek için gerçekleştirilen PCAR analizi sonucunda elde edilen bulgular, testin Rasch boyutundan farklı bir boyut içermediğini ortaya koymuştur. Bununla birlikte birbiri ile zıtlık oluşturan 1 ve 3. kümelerdeki maddeler dikkate alındığında göze çarpan husus olumlu ve olumsuz maddelerin ayrı ayrı kümelendikleridir. Oon ve Subramaniam (2011) tarafından gerçekleştirilen çalışmada da bu durumun aynısı ortaya çıkmış ve bunun neticesinden araştırmacılar gelecekte olumsuz maddelerin olumlu hâle çevrilerek uygulanmasını önermişlerdir. Psikolojik testlerde bireylerin maddeleri cevaplarken monotonlaşmasını önlemek, dikkatlerini üst seviyede tutmak, aynı desende yanıt vermelerinin önüne geçmek gibi nedenlerden ötürü olumsuz maddelere yer verilmektedir (Zeng, Wen ve Zhang, 2020). Burada yapılan kabul bir olumsuz madde ters sırada puanlandığında ölçme açısından olumlu hâliyle aynı işlevi göstereceğidir. Lakin bu hususta yapılan çalışmalar bu varsayımın geçerli olmayabileceğine işaret etmektedir. Örneğin, Carlson vd. (2011) olumsuz maddelerin nasıl işlev gösterdiğini inceledikleri araştırmada, olumsuz maddelerin testten elde edilen toplam puan ile zayıf ilişki gösterdiğini ve olumlu maddelere kıyasla daha düşük iç tutarlılığa sahip olduklarını ortaya koymuşlardır. Yine bu amaçla gerçekleştirilen bir başka çalışmada Suárez-Alvarez vd. (2018), olumlu ve olumsuz maddelerin bir arada yer almasının puanların varyansını azalttığını ve testin tek boyutlu olma durumunun ikincil varyans kaynaklarından ötürü bozulduğunu ifade etmektedir. Benzer başka bir çalışmada Zeng vd. (2020), bu iki madde türünün bir arada bulunmasının iç tutarlılığı düşürdüğünü ve faktör yapısını değiştirdiğini saptamışlardır. Tüm bunlar dikkate alındığında bu çalışma da gerçekleştirilen PCAR analizi sonucunda Rasch boyutundan kaynaklı toplam varyansın literatürde belirtilen eşik değerin biraz altında çıkmasının testin içerisinde olumlu ve olumsuz maddelerin bulunmasından ileri geldiği düşünülmektedir. Boone vd. (2014) kendi ölçme deneyimleri temelinde testlerde olumsuz maddelerin amaçlandığı şekilde işlemediğini ve dolayısıyla bu maddelere yer vermekten kaçınılmasını önermektedir. Tüm bu ifadelerden hareketle ileriki araştırmalarda olumsuz maddelerin olumlu forma dönüştürülerek kullanılması değerlendirilebilir.

Araştırmada ele alınan bir diğer husus, testte yer alan maddelerin cinsiyet ve sınıf düzeyi açısından işlev farklılaşması sergileyip sergilemediğiydi. Bu demografik değişkenler açısından gerçekleştirilen MİFA sonucunda maddelerin herhangi bir yanlılık sergilemediği saptanmıştır. Sonuç olarak test içerisindeki maddeler, bu değişkenler açısından bireylerin kaygı seviyeleri arasında farklılığı araştırılan çalışmalarda kullanılabilir.

Psikolojik değişkenleri incelemek için tasarlanan testlerin işlevsel açıdan geliştirilmesi uzun soluklu bir süreçtir. Bu amaçla ortaya konmuş bir testin ilk hâlinin kusursuza yakın şekilde işlev sergilemesi çok küçük bir ihtimaldir (Boone ve Staver, 2020). Buradan hareketle, bu çalışmada yapılan inceleme sonucunda ileri sürülen önerilerin ileriki çalışmalarda farklı örneklemeler üzerinde test edilmesi gerekmektedir. Böylece ilkökul seviyesinde öğrenim görmekte olan öğrencilerin matematik kaygı seviyelerine ilişkin daha geçerli ve güvenilir bir ölçüm gerçekleştirilmesi mümkün olacaktır.

Araştırmacıların Katkı Oranı Beyanı

Sorumlu yazar %40, ikinci yazar %30, üçüncü yazar %30.

Destek ve Teşekkür Beyanı

Trabzon Milli Eğitim Müdürlüğü'ne sağladıkları destekten dolayı teşekkür ederiz.

Çıkar Çatışması Beyanı

Çalışmada herhangi bir çıkar çatışması bulunmamaktadır.

KAYNAKLAR

- 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(2), 347-365.
- Bekdemir, M. (2009). Meslek yüksekokulu öğrencilerinin matematik kaygı düzeylerinin ve başarılarının değerlendirilmesi. *Erzincan Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 2(2), 169-189.
- Bond, T. G. ve Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*. New Jersey: Lawrence Erlbaum Associates.
- Bond, T. G., Yan, Z. ve Heene, M. (2021). *Applying the Rasch model: Fundamental measurement in the human sciences*. New York: NY: Routledge.
- Boone, W. J. (2020). Rasch basics for the novice. M. S. Khine (Ed.). *İçinde Rasch measurement : Applications in quantitative educational research* (s. 9-30). Singapore : Springer.
- Boone, W. J., Staver, J. R. ve Yale, M. S. (2014). *Rasch analysis in the human sciences*. Dordrecht: Springer.
- Boone, W. J. ve Staver, J. R. (2020). *Advances in Rasch analysis in the human sciences*. Cham, Switzerland: Springer.
- Brentari, E. ve Silvia, G. (2007). Unidimensionality in the Rasch model : How to detect and interpret. *Statistica*, 67(3), 253-261.
- Büyüköztürk, Ş., Kılıç-Çakmak, E., Akgün, Ö., Karadeniz, Ş. ve Demirel, F. (2019). *Bilimsel araştırma yöntemleri*. Ankara: Pegem Akademi Yayıncılık.
- 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. *Psychological Assessment*, 23(2), 582-562. doi: 10.1037/a0022484.
- Çepni, S. (2007). *Araştırma ve proje çalışmalarına giriş*. Trabzon: Celepler Matbaacılık.
- Duncan, P. W., Bode, R. K., Lai, S. M. ve 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. doi:10.1016/S0003-9993(03)00035-2
- Linacre, J. M. (1998). Structure in Rasch residuals: Why principal components analysis (PCA)? *Rasch Measurement Transactions*, 12(2), 636.
- Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3(1), 85-106.
- Linacre, J. M. (2022). A user's guide to Winsteps Ministeps Rasch model computer programs. Erişim adresi : <https://www.winsteps.com/a/Winsteps-Manual.pdf>
- Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for Research in Mathematics Education*, 30(5), 520-540.
- Manav, F. (2011). Kaygı kavramı. *Toplum Bilimleri Dergisi*, 5(9), 201-211.
- Mutlu, Y. ve Söylemez, İ. (2018). İlkokul 3. ve 4. sınıf çocukları için matematik kaygı ölçeği: Güvenirlilik ve geçerlik çalışması. *Ekev Akademi Dergisi*, 22(73), 429-440.
- Mutlu, Y., Söylemez, İ. ve Yasul, A. F. (2017). İlkokul öğrencilerinin matematik kaygısı ile matematik başarıları arasındaki ilişkinin incelenmesi. *Journal of Human Sciences*, 14(4), 4425-4434.
- Oon, P. T. ve Subramaniam, R. (2011). Rasch modelling of a scale that explores the take-up of physics among school students from the perspective of teachers. R. F. Cavanagh, ve R. F. Waugh (Ed.). *İçinde Applications of Rasch Measurement in Learning Environments Research* (s. 119-139). Rotterdam: Sense Pub.
- Smith Jr, E. V., Conrad, K. M., Chang, K. ve Piazza, J. (2002). An introduction to Rasch measurement for scale development and person assessment. *Journal of Nursing Measurement*, 10(3), 189-206.

- Suárez-Alvarez, J., Pedrosa, I., Lozano, L. M., García-Cueto, E., Cuesta, M. ve Muñiz, J. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema*, 30(2), 149-158. doi: 10.7334/psicothema2018.33
- Türk, C. ve Bedir, G. (2021). İlkokul 4. sınıf öğrencilerinin matematik kaygısı ile ders başarısı arasındaki ilişki. *Journal of Educational Reflections*, 5(2), 73-88.
- Wright, B. D. ve Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8(3), 370.
- Wright, B. D. ve Masters, G. N. (1990). Computation of outfit and infit statistics, *Rasch Measurement Transactions*, 3(4), 84.
- Wright, B. D. ve Stone, M. H. (1979). *Best test design*. Chicago : Mesa Press.
- Wright, B.D. ve Stone, M.H. (2004). *Making measures*. Chicago : Phaneron Press.
- Yenilmez, K. ve Özbey, N. (2006). Özel okul ve devlet okulu öğrencilerinin matematik kaygı düzeyleri üzerine bir araştırma. *Uludağ Üniversitesi Eğitim Fakültesi Dergisi*, 19(2), 431-448.
- Zeng, B., Wen, H. ve 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.

EXTENDED ABSTRACT

In addition to cognitive abilities, affective qualities towards mathematics and towards components related to mathematical learning environments have an important role in learning mathematics. In the affective context, one of the factors that play a role in the process of learning mathematics is mathematics anxiety. Due to the belief that students studying at primary school level may not have mathematics anxiety, there is less study that focused on this educational level compared to others (Mutlu & Söylemez, 2018). However, in the study conducted by Mutlu et al. (2017) it was determined that students studying at the 3rd and 4th grades of primary school have high levels of mathematics anxiety. In another study conducted at the primary education level, Yenilmez and Özbey (2006) found that students in lower grades had relatively higher levels of anxiety than students in upper grades. As it is understood from the literature, mathematics anxiety can be experienced by students at every educational level. Moreover, it also has an important role in understanding topics addressed in mathematics lessons. Therefore, necessary support should be given to students who experience mathematics anxiety in order to prevent anxiety-related failures in mathematics lessons. To provide such support, the first step is to identify students who experience mathematics anxiety. For this purpose, a survey study was carried out with 3rd and 4th grades students in which a mathematics anxiety scale developed by Mutlu and Söylemez (2018) was used. The aim in this study was to investigate the psychometric properties of the test within the framework of Rasch Measurement Theory. To this aim, the research questions addressed in the study are i) Does the category structure adopted to answer the items in the scale function well? ii) What are the person-item reliability and separation indices? iii) What are the fit indices of the items? iv) Does the scale target the individuals it aims to measure well? v) Is the distribution of the items appropriate? vi) Are there any biased items? vii) Does the scale meet the unidimensionality criterion?

To assess the functioning of the category structure, the five criteria proposed by Lincare (2002) were investigated. The findings showed that the distance between the thresholds did not meet the suggested criteria. Moreover, the category thresholds were not monotonically increasing. As to the reliability, the reliability index of the raw scores was calculated as .86, the reliability index of person measures in logit units was found to be .72, and the reliability index of item measures in logit units was calculated as .99. These figures indicate that the reliability indices are greater than .7, which is the lower limit recommended in the literature (Bond et al., 2021). In the study, person separation index was 1.6 and item separation index was 20.96. Considering the intervals proposed by Duncan et al. (2003) the person separation index is within the acceptable range. Based on the interpretation of Boone et al. (2014) the item separation index indicates that the data is appropriate to conduct analysis both at the individual and group levels. To evaluate to what extent do the items fit the model, the outfit and infit indices for item calibrations were calculated. The indices obtained were found to be within the acceptable range recommended in the literature. In addition to this, point-measure correlations of the items were all positive.

One of the factors that increase the content validity of a scale is that the items in the scale can tap different levels of the construct that is aimed to be measured (Boone et al., 2014). To assess this, person-item map was constructed. The findings showed that the items were not confined to a short range, so the test is capable to cover different levels of the mathematics anxiety variable. It is a requirement that a test should target the sample that it intends to measure well. If a test targets the sample well then it is expected that the region where the items are concentrated should be sited near the region where the persons are concentrated. The spreads of the persons and the items as revealed in the person-item map showed that the test targeted the sample poorly. To test whether the measures of the test items differ with respect to

gender and grade level variables differential item functioning analyses were conducted. The result of the analyses showed that the measures of the items did not change significantly in the subgroups of the sample. To assess whether the test satisfy the unidimensionality criteria principal component analysis of the residuals was conducted. The result of the test revealed that the proportion of the total variance explained by the Rasch measures was %47 and the eigenvalue of the first contrast was 2.34. In addition to this, the disattenuated correlation coefficient between the first and third clusters was .99 which is greater than .7. Based on these figures it was concluded that the test met the unidimensionality criteria.

The findings regarding the functioning of the category structure implies that the exclusion of the undecided category and adopting four categories such as *Strongly agree-Agree-Disagree-Strongly disagree* may result in a more productive measurement. As to the targeting effectiveness of the test, the findings indicated that most of the items in the test were hard to agree with for the sample. Therefore, the items in the test do not serve enough to distinguish between the individuals in terms of mathematics anxiety. From this point of view, it is considered that it is necessary to enrich the scale with items that can tap lower levels of the mathematics anxiety.