

İlköğretim Matematik Öğretmeni Adaylarına Yönelik Matematik Dilinin Öğretimi Öz Yeterlik Ölçeğinin Geliştirilmesi

Erdem ÇEKMEZ^{1*}, Mustafa GÜLER¹, Taha Ertuğrul KUZU²

¹ Trabzon Üniversitesi, Fatih Eğitim Fakültesi, Matematik ve Fen Bilimleri Eğitimi Bölümü, Türkiye

² University of Education, Schwäbisch Gmünd, Almanya

Özet: Bu araştırmada ilköğretim matematik öğretmeni adaylarının matematik dilini öğretmeye ilişkin öz yeterliklerini belirlemek için bir ölçeğin geliştirilmesi amaçlanmıştır. Araştırmanın katılımcılarını 14 üniversitenin ilgili programının son yarıyılında öğrenim gören 419 öğretmen adayı oluşturmaktadır. Ölçeğin ne düzeyde geçerli ve güvenilir ölçme gerçekleştirdiği madde uyum istatistikleri, benimsenen kategori yapısının işlevselliği, kişilere ve maddelere yönelik güvenilirlik ve ayırt edicilik indeksleri, örneklemi hedefleme yeterliği, maddelerin oluşturduğu hiyerarşi, maddelerin yanlılık gösterme durumu, tek boyutluluk ve yerel bağımsızlık bileşenleri açısından incelenmiştir. Verilerin analizi, Rasch ölçeği teorisi çerçevesinde derecelendirme ölçeği modeli temelinde Winsteps® yazılımı aracılığıyla gerçekleştirilmiştir. Bulgular kategori yapısının işlevsel olduğunu, madde uyum istatistiklerinin literatürde önerilen aralıkta bulunduğunu, tek boyutluluk ve yerel bağımsızlık kısıtlarının sağlandığını, kişilere ve maddelere ilişkin güvenilirlik ve ayırt edicilik indekslerinin uygun olduğunu, maddelerin yanlılık göstermediğini ortaya koymuştur. Bununla birlikte ölçeğin örneklemi hedefleme hususunda orta düzey isabetsizlik sergilediği belirlenmiştir. Sonuç olarak ölçek makul düzeyde geçerli ve güvenilir bir ölçüm gerçekleştirmekte olup ileriki araştırmalarda kullanılabilir.

Makale Bilgileri

Araştırma makalesi

Gönderim Tarihi

27/06/2024

Kabul Tarihi

19/03/2025

Anahtar Kelimeler

Matematik dili,

Öz yeterlik,

Matematik

öğretmeni adayları.

1. Giriş

Matematiksel bilginin öğrenilmesinde dil önemli bir role sahiptir. Dil, öğretmenlerin matematiksel kavramları ve yöntemleri öğrencilere açıklamada, matematiksel metinleri anlamlandırmada ve problem çözüme gibi matematiksel aktivitelerde düşünceye aracılık etmektedir (Gökmen ve Ulusoy, 2022). Bununla birlikte, öğrencilerin hem akranlarıyla hem de öğretmeni ile fikir alışverişinde bulunabilmesine olanak sağlamaktadır. Dilin aracılık ettiği iletişimin matematik öğrenmede sahip olduğu öneme istinaden, ülkemiz ortaokul matematik dersi öğretim programında öğrencilerin dil ve sembollerini etkin biçimde kullanarak problem çözmesine, düşüncelerini sözlü olarak ifade etmelerine ve matematiksel temsiller aracılığı ile iletişim kurabilmesine fırsat tanınması vurgulanmaktadır (Millî Eğitim Bakanlığı [MEB], 2024a).

* Sorumlu Yazar: Erdem ÇEKMEZ E-mail: erdemcekmez@trabzon.edu.tr Adres: Trabzon Üniversitesi, Fatih Eğitim Fakültesi, 61300, Trabzon, Türkiye

The copyright of the published article belongs to its author under CC BY 4.0 license. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>

Matematik sınıflarında gerçekleşen öğretim sürecinde dil, genel anlamda iletişimin gerçekleşmesini mümkün kılan gündelik dil ile matematiksel bilginin temsiline ve iletişimine aracılık eden matematik dili olarak iki farklı biçimde ortaya çıkmaktadır. Matematik dili matematikçilerin ve öğrencilerin matematiksel fikirleri net biçimde ifade etmelerine olanak veren bir iletişim biçimidir (Moore, 1994). Matematik dilinin kullanımı matematiksel kavramların anlaşılması ve kavramlarla etkileşim kurulması için gereklidir (Açıl ve Zeybek, 2017; Erath vd., 2018; Prediger vd., 2019). Ayrıca bireylerin matematiksel ilişkileri tanımlamasına ve analiz etmesine, problem çözme sürecinde matematiksel akıl yürütmelerini somutlaştırmasına olanak tanır (Barwell, 2016).

Öğrencilerin ele alınan matematiksel içeriğe yönelik anlamlı öğrenmeyi gerçekleştirebilmesi için matematik diline ilişkin yeterliklerinin öğretim süreci boyunca sürekli gelişim göstermesi gerekmektedir. Beklenen bu gelişimin bir yönü, öğrencilerin alana özgü kelime ve terim dağarcıklarının öğretim kademeleri ile doğru orantılı olacak şekilde genişlemesidir. Matematiksel kelime dağarcığı öğrencilerin matematiksel fikirleri hem anlamasında hem de iletişim sürecinde etkili biçimde kullanmasında kilit rol oynamaktadır (Riccommini vd., 2015). Bu görüşün paralelinde Uptegrove (2015), matematik diline özgü sembolik kelimeleri anlamlı biçimde kavrayan öğrencilerin matematiksel bilgiye yönelik derin ve sağlam bir anlayış geliştirebileceğini ortaya koymuştur. Bununla birlikte matematik dili sadece matematiksel kelimeleri ve notasyonları öğrenmekten ibaret değildir. Matematiksel bilgiyi ifade ederken takınılan üslubu taklit edebilmek, matematiksel bilgiyi farklı gösterim alanlarında temsil edebilmek ve bireyin zihninde gerçekleşen düşünceleri kâğıt üzerinde somutlaştırarak matematiksel akıl yürütme sürecinin gerçekleşmesine aracı olmak gibi farklı eylemleri de kapsamaktadır (Moschkovich, 2015; Peng vd., 2020). Bu sebepten matematik dilinin iletişim rolünün yanı sıra, dilin öğrenmeye ve düşünmeye aracılık ettiği gerçeği göz önüne alındığında bilişsel bir fonksiyonu da bulunmaktadır (Erath vd., 2018; Kuzu, 2023; Kuzu ve Ratzke, 2024).

Matematiğin etkili öğretimi büyük ölçüde kuralları, kavramları, sembolleri ve terimleri kapsayan matematik dilinin doğru kullanımına bağlıdır (Gürefe, 2018). Bu dil gündelik hayattaki konuşma dilinden farklı olarak anlamı net biçimde belirli kelimelerin kullanımını, kendine özgü sözdizimi kurallarını ve konuşma üslubunu içermektedir. Örneğin, matematiksel kavramların tanımları teknik terimler, yoğun isim öbekleri, özel anlamları olan bağlaçlar ve örtük mantıksal ilişkiler barındırmaktadır (Schleppegrell, 2007). Tüm bu yönleri ile matematik dili matematiksel bilgiyi anlamada ve problemleri çözmede öğrenciler için zorluklar ortaya çıkarmaktadır. Matematiğe özgü olup günlük dilde bulunmayan sözcüklerin anlaşılması hem günlük dilde hem de matematik dilinde bulunan fakat farklı anlamları olan kelimelerin ayrımı, matematiğe özgü isim öbeklerinin kavranması bu zorluklardan bazılarıdır (Wilkinson, 2018). Bu husustaki zorlukların üstesinden gelemeyen öğrencilerin matematiksel bilgiye ilişkin yaptıkları açıklamaların yalnızca kendilerine anlamlı gelen biçimde gerçekleştiği belirtilmektedir (Moschkovich, 2015). Böyle bir durumda anlamın ortak kılınmamasından ötürü öğretmenin uygun dönütte bulunması güçleşecektir. Bu gerekçelerden ötürü, öğrencilerin öğrenmelerini desteklemek için öğretmenlerin matematik dilini kullanırken dikkatli olmaları ve bu konuda anlaşılır olmaları gerekir (Adler, 1999; Araujo vd., 2018). Bunun yanı sıra öğrencilerin yaptıkları sözel açıklamaların içerisinde şayet hatalı ifadeler varsa bunları düzelterek yeniden seslendirmeleri öğretmenler için sunulan bir diğer öneridir (Wilkinson, 2018). Her ne kadar bu öneriler dile getirilse de Moleko (2021), problem çözmeye ilişkin yapılan öğretimde öğretmenlerin dilsel faktörlere dikkat etmediğini, bunun da öğrencilerin problemleri etkili bir şekilde görselleştirmelerini ve çözmelerini zorlaştırdığını saptamıştır.

Matematik öğretmenlerinin öğretim sürecini öğrencilerin matematik diline yönelik yeterliklerini destekleyecek biçimde tasarlamaların gerektiği vurgulanıyor olsa da bunun vuku bulmadığı belirtilmektedir (Baber, 2011; Durkin ve Shire, 1991; Prediger, 2019). Gray (2004),

öğretmenlerin matematik sınıflarında matematik dilinin öğretimi amacıyla bir öğretim gerçekleştirip gerçekleştirilmesinde bu beceriyi öğretme hususundaki öz yeterliklerinin rol oynadığını söylemektedir. Bu bağlamda Gray, eğer bir öğretmenin matematik dilini öğretebileceğine ilişkin inancı düşükse bu yönde girişimde bulunmamasının kuvvetle muhtemel olduğunu iddia etmektedir. Buradan hareketle, matematik öğretmeni adaylarından gelecekteki öğretim süreçlerinde matematik dilinin öğretimine ilişkin takınacakları tutum, bu hususa ilişkin öz yeterliklerine bağlıdır. Matematik öğretmeni adaylarının bu husustaki öz yeterliklerini belirlemek için ise kabul edilebilir seviyede geçerli ve güvenilir ölçme araçlarına ihtiyaç vardır.

Matematik dili alanında değerlendirilebilecek ölçme araçlarından biri Kabael ve Yayan (2017) tarafından geliştirilmiş olup, matematik öğretmeni adaylarının matematik dilini öğretmeye ilişkin öz yeterliklerinden ziyade matematik dilini kullanmaya ilişkin kendi öz yeterliklerini ölçmeyi amaçlamaktadır. Örneğin, ilgili ölçekte yer alan maddelerden biri “Bir önermeyi matematik dilinde ifade edebilirim” şeklindedir. Dolayısıyla bu çalışmada oluşturulan ölçek bu yönüyle Kabael ve Yayan’ın ölçeğinden farklılaşmaktadır. Mevcut bir başka ölçek, Özgen ve Bindak (2008) tarafından geliştirilen matematik öğretmeni adaylarının matematik okuryazarlıklarına ilişkin öz yeterliklerini belirlemeyi amaçlayan ölçektir. İlgili ölçekte yer alan maddeler matematik dilinin öğretiminden ziyade, matematik dilinden daha geniş bir perspektifte öğretmen adaylarının kendi matematik okuryazarlıklarını değerlendirdikleri maddelerden oluşmaktadır. Bir diğer çalışmada Bali-Çalikoğlu (2002), matematik öğretmeni adaylarının matematik öğretiminde dile ilişkin görüşlerini belirlemek amacıyla bir ölçek geliştirmiştir. Ölçek içerisinde yer alan maddeler incelendiğinde, bu çalışmada ölçülmesi hedeflenen örtük değişken ile ilgili olmadığı değerlendirilmiştir. Gerçekleştirilen alanyazın taraması sonucunda bu çalışmada ölçülmesi hedeflenen örtük değişkene yönelik bir ölçme aracına rastlanmamıştır. Bu yönde bir ölçme aracının geliştirilmesi öğretmen adaylarının öz yeterliklerini belirlemenin yanı sıra, ölçek maddelerinin onaylanabilirlik düzeyine bağlı olarak öğretmen adaylarının kendilerini eksik hissettikleri alt yeterlikleri açığa çıkararak, hizmet öncesi dönemde öğretmen adaylarına nasıl bir destek sunulması gerektiği hususunda da fikir verecektir. Bu gerekçe ve önemden hareketle bu çalışmanın amacı, öğretmen adaylarının matematik dilinin öğretimine yönelik öz yeterliklerini ölçebilecek bir ölçeğin geliştirilerek psikometrik özelliklerinin Rasch Ölçme Teorisi (RÖT) bağlamında incelenmesidir.

2. Kavramsal Çerçeve

2.1. Matematik Dili

Matematik dili, günlük konuşma dilinin yardımıyla matematiksel bilginin somutlaştırılmasına ve paylaşımına aracılık eden bir iletişim biçimidir. Her ne kadar bir dil olarak anılsa da matematik dili insanların günlük iletişim amacıyla kullandıkları dilden bazı açılardan farklılaşmaktadır. Örneğin, ana dili matematik olan kimse yoktur. Bir başka açıdan, bireyler konuşma dilini okur-yazar olmadan akıcı biçimde konuşabilirken matematik dilinde akıcı bir şekilde konuşabilmek için öncelikle matematik okur-yazar olmak gereklidir. Diğer taraftan matematik dilinin işlevini yerine getirebilmesi için bir ana dile ihtiyacı vardır. Bunlara ek olarak Burns (2006), matematik dilinin öğrenilmesinin bir yabancı dilin öğrenilmesine benzediğini belirtmekle birlikte matematik dilini öğrenmek için ilk olarak üzerinde konuşulacak kavramların öğrenilmesinin gerekli olduğunu vurgulayarak iki durumun farkına dikkat çekmektedir. Ayrıca, matematik dilini günlük konuşma dilinden farklılaştıran en önemli ayırt edici karakteristiği kesinlik içermesi ve belirsizlikten arınık olmasıdır. Bu karakteristik özellikler matematik dilini kusursuz bir muhakeme gerçekleştirme ve teknik terimleri tanımlama hususlarında faydalı bir araç kılmaktadır (Baber, 2011).

Bir kavram olarak matematik dili ifadesi matematik eğitimi alanyaznında çokça kullanılmasına rağmen üzerinde uzlaşılmış bir tanımı bulunmamaktadır. Örneğin Barton’un (2008) kaleme

aldığı “The language of mathematics” isimli kitap içerisinde matematik diline ilişkin açık bir tanıma rastlanmamaktadır. NCTM (2000) matematik dilinin etkin biçimde kullanılmasını matematiksel okuryazarlık terimi çerçevesinde ifade ederken açık bir tanım vermemektedir. Ülkemizde 2018 yılında yürürlüğe girmiş ilkökul ve ortaokul matematik dersi öğretim programında (MEB, 2018a) matematik dili ifadesi açıkça yer almakta fakat bir tanım verilmezken, ortaöğretim matematik dersi programında (MEB, 2018b) ise matematik dili ifadesine rastlanmamıştır. 2024 yılında yürürlüğe girmiş ortaokul programında (MEB, 2024b) ise matematik dili ifadesi yer almazken, ortaöğretim programında (MEB, 2024a) matematiksel dil ifadesi bulunmakta olup 2018 programında olduğu gibi bir tanıma rastlanmamıştır. Dolayısıyla, matematik dili ifadesinden üstü örtük olarak herkesin aynı şeyi anlayacağına dair bir genel kabulün olduğu değerlendirilmektedir.

Matematik eğitimi alanyazını incelendiğinde farklı araştırmacıların matematik dilini açıklamaya veya tanımlamaya ilişkin girişimlerinin olduğu görülmektedir. Bu bağlamda Baber (2011), matematiğin dilini matematik yaparken kullanılan notasyon biçimleri olarak ifade etmekte ve matematik dilinde yazılan metinlerin değerler, değişkenler, fonksiyonlar ve ifadeler olmak üzere 4 temel öğeden oluştuğunu söylemektedir. Halliday (1978) ise matematik dilini bir üslup olarak ele alıp, günlük konuşma dilinin matematiksel amaçlar için kullanılması olarak tanımlamakta ve sembolizmi matematiğin diline dâhil etmemektedir. Diğer taraftan Morgan (2020), matematik dilini tarif ederken karakteristik özelliklerini sıralamıştır. Bu karakteristik özellikleri ise alana özgü kelimeler içermesi, fiillerden matematiksel terimler oluşturma özelliğinin olması –öteleme teriminin türetilmesi gibi–, cebirsel ifadeler barındırması ve Kartezyen koordinat sistemi gibi farklı semiyotik sistemlerden yararlanması olarak ifade etmiştir. Morgan’a benzer biçimde Pimm (1987), matematik dilinin özelliklerini alana özgü anlamlar içermesi, semboller ile sembolize edilen varlıklar barındırması ve kendisine has sözdizimi (sentaks) kuralları içermesi olarak sıralamıştır.

Gray (2004), doktora çalışmasında matematik diline ilişkin açık bir tanımın olmadığına vurgu yaparak bir işlevsel tanım sunmayı amaçlamıştır. Bu amaç doğrultusunda yaptığı incelemenin sonucunda matematik dilini İngilizce dili bağlamında işlevsel anlamda “İngilizce dilinin konuşma dili olarak kullanıldığı sınıflarda matematik dili sembolik biçimlerden oluşan bir küme ile İngilizcenin bütünleştirilmesi ile kullanılır. Bu dil matematik hakkında tartışırken veya yazarken alana özgü kelimelere, sembollere, grafiklere ve teamüllere başvurur” (s. 21) olarak tanımlamıştır. Gray önerdiği bu tanıma matematik eğitimi alanında deneyimli araştırmacılar ile paylaştığını ve araştırmacıların tanımın uygun olduğu yönünde görüş bildirdiklerini ifade etmiştir.

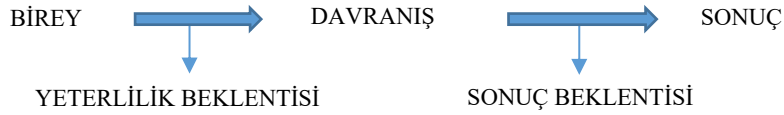
Matematik dilinin tanımına ilişkin yapılan alanyazın taramasının sonucunda matematik dilinin kendine has notasyonlarının, terimlerinin ve teamüllerinin olduğu; günlük konuşma dili ile kelimeler paylaştığı, farklı temsil ve semiyotik sistemlerden yararlandığı, alana özgü sözdizimi kurallarının bulunduğu, gelişiminde iletişimin önemli rol oynadığı anlaşılmaktadır. Bu çalışmada madde havuzunun oluşturulmasında bu nitelikler dikkate alınmıştır.

2.2. Öz Yeterlik

Öz yeterlik ilk olarak Albert Bandura tarafından ileri sürülmüş bir kavram olup, sergilenen davranışların bireylerin inanç ve duyguları ile bağlantılı olduğunu iddia etmektedir (Akcan, 2023). Bandura (1997) insan eylemliliğinin üçlü karşılıklı nedenselliği (triadic reciprocal causation) ihtiva eden birbirine bağlı nedensel bir yapı içerisinde işlediğini iddia etmektedir. Üçlü karşılıklı nedensellik yapısı davranış; bilişsel, duyuşsal veya biyolojik biçimde olan kişisel faktörler ve çevreden oluşmaktadır. Bu ilişki ağının öngörüsü bireylerin tercihlerinin yalnız çevresel faktörlerden değil, seçeceği bir tercihi gerçekleştirmeye ilişkin kabiliyetlerine yönelik kişisel değerlendirmelerine bağlı olduğudur.

İnsan eylemliliğinin belirleyicileri arasında olan öz yeterliği Bandura (1997, s. 21) “algılanan öz yeterlik, kişinin belirli performans türlerini organize etme ve yürütme yeteneğine ilişkin bir yargıdır” şeklinde tanımlamıştır. Bireyin bir problem ile karşılaştığında kendi öz yeterliğine yönelik inancı problemi çözmeye ilişkin bir girişimde bulunup bulunmayacağı, bu süreçte ne kadar çaba sarf edeceği ve karşılaşacağı olası zorlukların üstesinden gelmek için ne kadar metanet göstereceği hususlarında belirleyici olmaktadır (Bandura, 1977). Bandura (1997), bireylerin bir eylemin sonuçlarına ilişkin beklentilerinin büyük ölçüde o eylemi ne kadar iyi ifa edeceklerine yönelik değerlendirmelerine bağlı bulunduğunu belirtmektedir.

Şekil 1. *Yeterlik İle Sonuç Beklentilerinin Birey, Davranış ve Netice Silsilesindeki Konumu*



Bandura beklentiler hususunda bir ayırım olduğunu belirtmektedir. Şekil 1’de resmedildiği üzere, Bandura (1977) birey ile davranış arasında bireyin yeterlik beklentisinin, davranış ve sonuç arasında ise bireyin sonuç beklentisinin rol oynadığını söylemektedir. Sonuç beklentisi, gerçekleştirilen bir davranışın belirli bir sonuç ortaya çıkaracağına dair bireyin tahminini belirtmektedir. Yeterlik beklentisi ise bireyin ilgili sonucu ortaya çıkartacak davranışları gerçekleştirip gerçekleştiremeyeceğine dair kanaatidir. Bandura (1997), birey bir eylemler silsilesinin belirli sonuçları ortaya çıkartacağına inansa bile, gereken eylemleri gerçekleştirebilme hususunda şüpheye sahipse bu inancının davranışına etki etmeyeceğini söylemektedir.

Öz yeterlik kavramını tanıttığı kitabında Bandura (1997), kavramın öğretmenlerin öğretim faaliyetleri üzerindeki olası etkilerini de tartışmıştır. Bu hususta araştırmacı, elde edilen kanıtların öğretmenlerin yeterliklerine dair inançlarının sınıflarındaki akademik faaliyetlerini nasıl yapılandırdıklarını kısmen belirlediğine ve öğrencilerin entelektüel yeteneklerine ilişkin değerlendirmelerini şekillendirdiğine işaret ettiğini vurgulamaktadır. Buna örnek olarak, düşük düzeyde yeterlik algısına sahip öğretmenlerin, eğer öğrenciler düşük güdülenmeye sahipse kendilerinin yapabilecekleri çok az şey olduğuna ve öğrencilerin entelektüel gelişimi üzerinde uygulayabilecekleri etkinin çevreden kaynaklı olumsuz etkilerle ciddi şekilde sınırlandığına inanmalarını göstermektedir. Sonuç olarak, Bandura, öğrenmeyi teşvik etme becerilerine güçlü bir şekilde inanan öğretmenlerin, öğrencileri için olumlu öğrenme deneyimleri yaratacağını vurgulamaktadır. Buna karşılık, yeterlikleri konusunda kendilerinden şüphe duyan öğretmenlerin, öğrencilerin yeteneklerine ve bilişsel gelişimlerine ilişkin özgüvenlerini zedelemesi muhtemel sınıf ortamları oluşturacaklarını belirtmektedir.

Tüm bu söylenenler takip eden çıkarımı ortaya koymaktadır. Bir öğretmen matematik dilinin öğrencilere öğretilmesinin öğrencilerin öğrenme yaşantılarına olumlu sonuçlar getirebileceğine inanabilir (sonuç beklentisi), fakat bu beceriyi öğretebileceğine dair şüpheleri varsa (yeterlilik beklentisi) bu yönde bir eylem sergilemeyecektir. Dolayısıyla, öğretmenlerin bu beceriyi öğrencilerine kazandırabileceklerine ilişkin makul seviyede öz yeterliliğe sahip olması gerekmektedir.

2.3. Araştırmanın Problemi

Araştırmada ele alınan ana problem, matematik öğretmeni adaylarının gelecekteki meslek hayatlarında matematik dilini öğretmeye ilişkin sahip oldukları öz yeterlikleri belirlemek için geliştirilen ölçeğin psikometrik özelliklerinin neler olduğudur. Bu çerçevede araştırmada cevap aranacak alt problemler aşağıda listelenmiştir.

- Ölçekte benimsenen kategori yapısı uygun mudur?
- Kişi ve madde güvenirlik ile ayırt edicilik indeksleri önerilen eşiklerin üzerinde midir?
- Maddelerin uyum istatistikleri kabul edilebilir sınırlar içerisinde midir?
- Ölçek örnekleme iyi hedeflemekte midir?
- Maddeler nasıl bir dağılım sergilemektedir?
- Ölçek maddeleri yanlılık göstermekte midir?
- Ölçek Rasch boyutundan farklı bir boyut içermekte midir?
- Ölçek yerel bağımsızlık kriterini sağlamakta mıdır?

3. Yöntem

Var olan bir durumu araştırmak ve vasfına yönelik detaylı bilgiye ulaşmak için gerçekleştirilen incelemeler betimsel araştırmalar başlığı altında sınıflandırılmaktadır (Çepni, 2007). Bu başlık altında sınıflandırılan tarama yöntemi, kişilerde mevcut olan bir nitelik hakkında bilgi sahibi olmak için veri toplanmasını amaçlamaktadır (Büyüköztürk vd., 2019). Tarama yöntemi ile gerçekleştirilen bu araştırmada, öğretmen adaylarının gelecekteki meslek hayatlarında matematik dilinin öğretimine ilişkin sahip oldukları öz yeterlikleri belirlemek için geliştirilen ölçeğin psikometrik özellikleri incelenmiştir. Çalışma için etik onay, Trabzon Üniversitesi Sosyal ve Beşeri Bilimler Bilimsel Araştırma ve Yayın Etiği Kurulu'ndan alınmıştır.

3.1. Katılımcılar

Araştırmanın katılımcılarını [Tablo 1](#)'de belirtilen üniversitelerde ilköğretim matematik öğretmenliği programının son yarıyılında öğrenim gören öğretmen adayları oluşturmaktadır. Bu üniversitelerin seçiminde farklı bölgelerde yer alması ve taban puanları açısından farklılık göstermesine dikkat edilmiştir. Böylece örneklem çeşitliliğinin artırılması amaçlanmıştır. Ölçeğin hedeflediği niteliğin gelişiminde öğrencilerin fakülte hayatlarında aldıkları farklı derslerin katkısı olduğu değerlendirilmektedir. Bu sebepten son yarıyıda öğrenim gören öğrencilerin örneklem olarak alınmasının uygun olacağı değerlendirilmiştir.

Seçer (2018), maddelerin ölçekle uyumunu incelemek amacıyla gerçekleştirilecek bir pilot uygulamada eğer madde sayısı 30'un altında ise 50 kişilik bir grubun yeterli olduğu yönünde görüş bildirmektedir. Ölçek geliştirme sürecinin esas uygulama safhasında ise Tabachnick ve Fidell (2013) en az 300 sayısını, Nunnally (1978) ise madde sayısının 10 katını önermektedir. Ölçeğin pilot uygulaması 68 öğrenci ile gerçekleştirilmiştir. Esas uygulamada ise ölçeği 457 öğretmen adayı cevaplamıştır. Yapılan ilk incelemede “Bu maddede en alt sırada bulunan yorum yok seçeneğini işaretleyiniz” ifadesiyle sunulmuş kontrol maddesine “yorum yok” haricinde yanıt veren 33 kişi olduğu tespit edilmiş ve bu kişiler veri setinden çıkartılmıştır. Bununla birlikte ölçekten tam puan alan 5 kişi de veri setinden çıkartılmıştır. Analize dâhil edilen 419 kişinin kategorik değişkenlere göre dağılımı [Tablo 1](#)'de sunulmuştur.

Tablo 1. Örneklemin Kategorik Değişkenlere Göre Dağılımı

	Üniversite														Cinsiyet	
	Trabzon	Bayburt	İstanbul Medipol	Yozgat Bozok	Recep Tayyip Erdoğan	Sakarya	Sivas Cumhuriyet	Dokuz Eylül	Düzce	Manisa Celal Bayar	Necmettin Erbakan	Giresun	Erzincan Binali Yıldırım	Kilis 7 Aralık	Kadın	Erkek
n	62	28	16	21	35	25	24	49	38	10	29	29	17	36	330	89

3.2. Ölçek Formunun Geliştirilme Süreci

Çalışmada ilk olarak maddelerinin yazımında dayanak noktası oluşturmak için matematik dilinin nasıl tanımlandığına ve neleri ihtiva ettiğine yönelik alanyazın taraması yapılmıştır. Buradan elde edilen bulgular temelinde madde havuzu oluşturulmuştur. Alanyazında olumsuz maddelerin arzu edilen şekilde işlemediğini rapor eden araştırmalar (Boone vd., 2014; Carlson vd., 2011; Suárez-Alvarez vd., 2018; Zeng vd., 2020) dikkate alınarak madde havuzunda olumsuz maddeye yer verilmemiştir.

Madde havuzu oluşturulduktan sonra ilk olarak 9 öğretmen adayı ile maddeler üzerinde grup görüşmesi gerçekleştirilmiştir. Bu görüşmede araştırmacılardan biri her bir maddeyi öğretmen adaylarıyla birlikte tek tek okumuş ve maddenin anlaşılabilir olup olmadığı ile ne anlam ifade ettiği hakkında fikir beyan etmelerini istemiştir. Gelen geri bildirimler temelinde bazı maddelerde değişiklik gerçekleştirilmiştir. Bu işlemin akabinde ölçeğin farklı açılardan geçerliliğini arttırmak için ölçek maddeleri 13 farklı üniversitede matematik eğitimi alanında görev yapan 18 öğretim elemanı (6 Profesör, 11 Doçent, 1 Doktor Öğretim Üyesi) ile çevrimiçi ortamda paylaşılmış ve maddeler hakkında değerlendirme yapmaları istenmiştir. Geri dönütler dikkate alınarak gerçekleştirilen iyileştirmeler neticesinde taslak ölçek pilot uygulamaya hazır hale getirilmiştir. Gerçekleştirilen pilot uygulama sonrasında son hali verilen ve 20 maddeden (Bkz. [Ek A](#)) oluşan ölçek esas uygulama için hazır hale getirilmiştir.

3.3. Ölçeğin Uygulanma Süreci

Ölçek katılımcılara Google Formlar platformunda çevrimiçi olarak gönderilmiştir. Ölçeğe ulaşmak için gönderilen bağlantıya tıklandığında katılımcılar bir davet mektubu ile araştırmaya gönüllü olup olmadıklarını onaylamalarını isteyen bir seçenek ile karşılaşmışlardır. Gönüllü olmayı tercih eden katılımcılara, cinsiyet ve öğrenim gördükleri üniversiteyi belirttikten sonra ölçeğe ilişkin bir açıklama metni sunulmuştur.

Şekil 2. Ölçek Maddeleri İle Birlikte Verilen Açıklamalar

Matematiksel Notasyon: Matematiksel kavramları temsil eden biçimlerdir. Örneğin:

$$\int f(x)dx, \frac{a}{b}, m_d \text{ vb.}$$

Matematiksel Sembol: Matematiksel ilişkileri ya da işlemleri göstermek için kullanılan biçimlerdir. Örneğin: $\leq, \neq, \div$ vb.

a)

Matematiksel Terim: Matematiğe özgü olan sözcüklerdir. Örneğin: payda, çarpma işlemine göre tersi, açığortay, giriş, kalan, bölüm, onda birler basamağı vb.

b)

Ölçekte yer alan maddeler için katılımcıların yalnız birini onaylayabilecekleri *çok az yeterliyim, az yeterliyim, orta düzeyde yeterliyim, yeterliyim, çok yeterliyim* şeklinde 5 kategori bulunmaktadır. Matematiksel terim veya matematiksel notasyon/sembol ifadelerini içeren maddeler sırasıyla [Şekil 2a](#) ve [Şekil 2b](#)'de görülen açıklamalar ile birlikte sunulmuştur.

Katılımcıların ölçek maddelerini yanıtlarken rasgele işaretleme yapıp yapmadıklarını belirlemek için onuncu sırada yer alan bir kontrol maddesi konulmuştur. Bu kontrol maddesinde katılımcılardan yalnızca bu madde de yer verilen *yorum yok* seçeneğini onaylamaları istenmiştir. Son olarak, katılımcıların yanıtları puanlanırken yüksek puanın yüksek öz yeterliği işaret ettiği kurgusu benimsenmiştir.

3.4. Veri Analizi

3.4.1. Derecelendirme Ölçeği Modeli

Psikolojik yapıların ölçülmesi amacıyla derecelendirme ölçekleri yaygın olarak kullanılmaktadır. Lakin, bu ölçeklerden elde edilen ham puanların araştırma sorularına cevap bulmak için doğrudan istatistiksel analizlere tabi tutulması birçok sınırlılığı beraberinde getirmektedir (Boone vd., 2014). Bu sınırlılıklardan biri, elde edilen verilerin parametrik testlerin varsayımlarından biri olan normallik şartını sağlamayabileceğidir. Böyle bir senaryoda puanların ortalamaları üzerinden gerçekleştirilecek yorumlar yanıltıcı olabilmektedir (Neumann vd., 2011).

Diğer taraftan, bu türden ölçeklerde katılımcılara onaylayabilecekleri 2'den fazla sayıda kategori sunulmakta ve bu kategorilere verilen yanıtlar ardışık tam sayılar ile puanlanmaktadır. Bu pratik ise ardışık kategorilerin, ölçülmesi hedeflenen örtük değişkenin miktarı açısından eşit farklılığa işaret ettiği kabulünü beraberinde getirmektedir. Bununla birlikte herhangi iki maddede aynı kategori onaylanmışsa, her ikisinden de eşit puan elde edilmesi ve bu ham puanların toplam puana katkısının aynı olmasından ötürü örtük değişkenin miktarına eşit katkı yaptığı kabulü benimsenmektedir. Bu kabullerin temelinde ise ölçekten elde edilen verilerin eşit aralıklı olduğu varsayımı bulunmaktadır. RÖT'ü benimseyen araştırmacılar yaygın uygulamaya sahip bu pratiğin hatalı olduğunu savunmakta hem kategorilerin hem de maddelerin örtük değişkenin miktarına yönelik farklı katkı sunabileceğini değerlendirmektedir (Boone vd., 2014). Bir başka ifadeyle, bir ölçekten 20 ile 30 ham puan elde eden iki bireyin arasındaki farkın, 30 ile 40 puan elde eden iki birey arasındaki farka eşit olmayabileceğini belirtmektedirler. Dolayısıyla, derecelendirme ölçeklerinden elde edilen ham puanlara sıralama tipinden veri olarak muamele edilmesi gerektiğini savunmaktadırlar (Bond ve Fox, 2007).

Ölçek geliştirme sürecinde ve ölçeklere verilen yanıtların sayısallaştırılmasında Rasch modelleri önemli avantajlar sunmaktadır (Bond ve Fox, 2007). Bu avantajlardan biri elde edilen ham puanların sıralama türünden veri olarak kabul edilerek, bu puanlardan eşit aralıklı puanlar oluşturulmasıdır (Boone vd., 2014). Bunun yanı sıra, Smith ve diğerleri (2002), Klasik Test Teorisine (KTT) kıyasla RÖT'ün sahip olduğu diğer bazı avantajları şu şekilde sıralamaktadır: (i) veri kaybı problemine çözüm sunması, (ii) madde ölçülerinin yanı sıra kişi ölçüleri için de geçerlik ve güvenirlik kestirimleri sağlaması, (iii) grup istatistiklerinden ziyade tekil olarak kişileri ve maddeleri odağına alması, (iv) kişilerin ve maddelerin ölçülerinin eş ölçü birimi cinsinden eşit aralıklı bir ölçek üzerinde konumlandırılabilmesi. Gerçekleştirilen bir analiz sonucunda maddelerin ve kişilerin ölçülerinin logit olarak isimlendirilen aynı ölçü birimi cinsinden ifade ediliyor oluşu, maddelerin ve kişilerin hem kendi içerisinde hem de birbiriyle mukayesesi olanaklı kılmaktadır. Bu ise özellikle eğitimcilere yaptıkları bir ölçüm sonucunda daha detaylı yorum yapabilecekleri bilgi kümesi sunmaktadır.

RÖT bünyesinde derecelendirme tipi ölçeklerin analizinde Derecelendirme Ölçeği Modeli'nden[†] (Rating Scale Model [DÖM]) yararlanılmaktadır. DÖM'ün işleyişi aşağıda görülen matematiksel denkleme dayanmaktadır. Denklem içerisindeki F_j değişkeni peş peşe gelen iki kategorinin seçilme olasılıklarının eşit olduğu eşik değerleri göstermek olup, tanımından anlaşılacağı üzere kategori sayısının bir eksiği kadar farklı eşik değeri bulunmaktadır. Bu araştırma özelinde örneklendirilecek olursa, F_4 değeri “çok yeterliyim” ile “yeterliyim” kategorilerinin seçilme olasılığının eşit olduğu eşik değeri göstermektedir. Diğer değişkenler arasında yer alan B_n n'inci kişinin ölçüsünü, D_i i'inci maddenin ölçüsünü ve son olarak P_{nij} ile $P_{ni(j-1)}$ ise ardışık kategorileri seçme olasılıklarını temsil etmektedir.

[†] Modelin işleyişine ilişkin açıklamaya Çekmez ve diğerlerinin (2023) araştırma raporundan ulaşılabilir.

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

3.4.2. Kategori Yapısının İşlevselliği

Derecelendirme türünden ölçekler ile gerçekleştirilen bir ölçme işleminde elde edilen ölçümlerin geçerliliği ilk olarak ölçekte benimsenen kategori yapısının araştırmacıların işlemesi istediği şekilde işlev gösterip göstermediğine bağlıdır. Çünkü ölçümden elde edilen kestirimler bu kategorilere verilen yanıtların sayısallaştırılması sonucunda elde edilmektedir. Tasarlanan bir ölçekte benimsenen kategori yapısının olması gerektiği gibi işlediğinin bir varsayımdan ziyade sınanması gereken bir durum olarak kabul edilmesi gerekmektedir (Boone vd., 2014). Bu yaklaşımın gerekçesi, yanıtlayıcıların maddelere gösterdikleri tepkilerin testi tasarlayanların beklentilerinden farklılık gösterebilmesidir (Andrich, 1996).

Katılımcıların maddelere verdikleri yanıtlar *Çok az yeterliyim* kategorisinden *Çok yeterliyim* kategorisine doğru artan şekilde 1-5 arası tamsayılar ile puanlanmıştır. Bunun doğal sonucu olarak, herhangi bir maddede herhangi bir kategoriye onaylayan katılımcıların ölçülerinin ortalaması, daha düşük puanlı bir kategoriye onaylayan katılımcıların ölçülerinin ortalamasından daha büyük olmalıdır. Bunun yanı sıra, kategorilerin işlevsel olup olmadığını sınamak için Linacre (2002) şu altı kıstası ileri sürmüştür: (i) kategori ortalama ölçüleri monoton artmalıdır, (ii) tüm kategoriler en az 10 kez onaylanmalıdır, (iii) dış uyum istatistikleri 2'den büyük olmamalıdır, (iv) kategori eşik ölçüleri monoton artmalıdır, (v) ardışık eşikler arasındaki fark 1.4 logit değerinden büyük fakat 5 logit değerinden küçük olmalıdır, (vi) kategori olasılık eğrilerinin her biri zirve yapmalıdır. Kategorilerin işlevselliğini değerlendirmede bu hususlar dikkate alınmıştır.

3.4.3. Maddelerinin Dağılımı ve Ölçeğin Hedefleme Yeterliği

Bir ölçeğin yapı geçerliliğini etkileyen faktörlerden biri ölçmeyi amaçladığı niteliğin farklı düzeylerine karşılık gelen maddeleri barındırmasıdır (Boone vd., 2014). Maddelerin ölçüleri dar bir aralığa hapsolmuşsa, maddeler ölçülecek niteliğin düşük, orta ve yüksek seviyelerini belirlemede yetersiz kalacaktır. Rasch analizi içinde ölçek maddelerinin hiyerarşisini ve ölçülecek niteliğin farklı seviyelerine hitap edip etmediğini değerlendirmede kullanılan araçlardan biri kişi-madde (Wright-map) haritasıdır. Kişi-madde haritası, logit olarak isimlendirilen ölçme birimi cinsinden kişileri ve maddeleri eş aralıklı bir ölçek üzerinde konumlandırır. Bu haritanın bir diğer işlevi ise, ölçeğin uygulandığı örneklemin hedefleme yeterliliğini değerlendirme imkânı sunmasıdır (Boone vd., 2014).

Tasarlanmış bir ölçeğin uygulandığı örneklemin hedefleme yeterliği, geçerliğini etkileyen faktörlerden bir diğeridir. Bir ölçeğin örneklemin iyi hedeflediğinin anlamı, örneklemdaki bireyler için onaylanabilirlik açısından ne çok zor ne de çok kolay olmasıdır (Boone vd., 2014). Ölçeğin örneklemin iyi hedeflediğinin göstergelerinden biri, kişi-madde haritasında kişilerin yoğunlaştığı bölge ile maddelerin yoğunlaştığı bölgenin karşılıklı olmasıdır (Bond ve Fox, 2007). Böyle bir durumda kişiler için yapılacak kestirimlerde hata azalacak ve kişileri birbirinden daha hassas biçimde ayırt etmek mümkün olacaktır. Kişilerin ortalaması ile maddelerin ortalaması arasındaki farkın mutlak değerinin 0.5'ten küçük olması hafif isabetsizliğe, 1'den büyük olması ise kayda değer bir isabetsizliğe işaret etmektedir (Duncan vd., 2003). Ölçeğin maddelerinin dağılımının uygunluğu ve hedefleme yeterliğini değerlendirmede bu hususlar dikkate alınmıştır.

3.4.4. Madde Uyum İstatistikleri

RÖT bağlamında uyum ölçme işlemi sonucunda elde edilen verilerin modele ne düzeyde benzerlik sergilediğinin göstergesidir (Bond vd., 2021). Modelin beklentisi ile yüksek düzeyde uyumsuzluk sergileyen maddeler ölçeğin hedeflediği niteliğin ölçümüne hizmet etmemektedir (Oon ve Subramaniam, 2011).

Uyum istatistiklerinin raporlaştırılması artık kareler ortalaması (Mean-square residuals) ile gerçekleşmektedir. Bu istatistiklerden birincisi olan iç uyum (infit) bireylerin kendi ölçülerine yakın maddelere verdikleri beklenmedik yanıtlara hassas iken, diğer uyum indeksi olan dış uyum (outfit) bireylerin kendi yetenek ölçülerinden uzakta olan maddelere verdikleri beklenmedik yanıtlara hassastır (Bond ve Fox, 2007). Gerçekleştirilen bir Rasch analizi sonucunda hem kişilere hem de maddelere ilişkin uyum indeksleri elde edilmektedir. Bununla birlikte, bir ölçekte yer alan maddelerin ölçülerinin değişmezlik sergilemesi istendiğinden, madde uyumu kişi uyumuna nazaran daha ön planda yer almaktadır (Wright ve Linacre, 1994). Wright ve Linacre (1994), derecelendirme ölçeklerine yönelik uyum indekslerinin 0.6-1.4 aralığında olmasını önermektedir.

Bir ölçekte yer alan maddelerin hedeflenen niteliğin ölçümüne hizmet edip etmediğine karar kılmada dikkate alınan bir diğer gösterge maddelerin nokta-ölçü korelasyonlarıdır (point-measure correlation) (Boone ve Staver, 2020). Bir maddenin korelasyon değerinin sıfıra yakın ya da negatif olması o maddenin ölçeğin bütünü ile zıt hareket ettiği anlamına gelmektedir (Bond vd., 2021). Maddelerin nokta-ölçü korelasyonuna ilişkin ileri sürülen alt eşik ise 0.3'tür (Boone, 2020). Maddelerin ölçeğin amacına hizmet edip etmediğini değerlendirmede bu hususlar da dikkate alınmıştır.

3.4.5. Güvenirlik ve Ayırt Edicilik

RÖT bağlamında güvenilirlik KTT'ye benzer biçimde yorumlanmaktadır (Linacre, 2022). Bu sebepten, katsayılar 1'e ne kadar yakınsa o denli iç tutarlılığı yüksek bir ölçme işleminin gerçekleştiği sonucuna varılır. KTT'den farklı olarak RÖT bağlamında hem kişilere hem de maddelere yönelik güvenilirlik katsayıları elde edilmektedir. Maddelere ilişkin güvenilirlik katsayısı, ölçülen nitelik temelinde ölçek maddelerinin bir hiyerarşi tanımlama potansiyeline işaret eder (Linacre, 2022). Kişi güvenilirlik indeksinin yüksek olması ise hem yüksek hem de düşük puan elde edilebilecek bir ölçek oluşturulduğu ve hesaplanan kişi ölçülerinin tutarlı olduğu anlamına gelir (Bond vd., 2021).

Rasch analizi sonucunda elde edilen ve güvenilirlikle doğrudan ilişkili olan bir diğer kavram ayırt edicilik indeksidir (G). Bu indeks gerçek puanların standart sapmasının hataların standart sapmasına bölünerek elde edilir (Linacre, 2022). Dolayısıyla örneklemin ölçme kesinliği temelinde dağılımını göstermektedir (Wright ve Stone, 2004). Kişiler için hesaplanan ayırt edicilik indeksi bireylerin ne düzeyde sınıflandırılabilirliğine ilişkin bilgi sunmaktadır (Bond vd., 2021). Kişi ayırt edicilik indeksinin 3'ten büyük olması mükemmel, 2-3 arasında olması iyi, 1.5 değeri ise kabul edilebilirlik sınırındır (Duncan vd., 2003). Madde ayırt edicilik indeksinin ise grup bazında analiz yapabilmek için en az 1.5, kişi bazında analiz gerçekleştirmek için ise 2.5 olması gereklidir (Tennant ve Conaghan, 2007). Gerçekleştirilen bir ölçme sonucunda örneklemin kaç farklı tabakaya parçalanabileceği kişi ayırt edicilik indeksi temelinde aşağıdaki formül ile hesaplanmaktadır (Linacre, 2022).

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

Ölçek maddelerinin sergilediği iç tutarlılık ve bireyleri ayırt etme gücüne ilişkin değerlendirmede bu hususlar dikkate alınmıştır.

3.4.6. Tek Boyutluluk ve Yerel Bağımsızlık

RÖT'ün temel varsayımlarından biri bir ölçme aracının yalnızca tek bir örtük değişkenin ölçümüne hizmet etmesi gerektiğini belirten tek boyutluluk (unidimensionality) ilkesidir (Linacre, 2022). Bir ölçme aracının bu ilkeyi sağlayıp sağlamadığını sınamada PCAR (Principal Component Analysis of Residuals) analizi kullanılmaktadır (Boone ve Staver, 2020). PCAR analizi sonucunda artıkların korelasyon matrisinden hesaplanan birinci temel bileşenin özdeğeri 2'den küçük ise, örtük değişkenin haricinde başka faktörlerden kaynaklı varyansın rasgele gürültü (random noise) olduğuna kanaat getirilmektedir (Boone ve Staver, 2020). Bunun haricinde, örtük değişkendeki değişimden kaynaklı varyans toplam varyansın yarısından fazlasını açıklıyorsa bu durumun da tek boyutluluğun başka bir göstergesi olduğu kabul edilmektedir.

RÖT'ün ikinci temel varsayımı yerel bağımsızlık ilkesidir (Bond vd., 2021). Bu ilke, bir maddeye verilen tepkilerin başka bir maddeye bağımlı olmaması gerektiğini ifade eder (Fan ve Bond, 2019). Şayet maddeler birbirinden bağımsız iseler, madde artıkları arasında yüksek korelasyon bulunmaması gerekir (Fan ve Bond, 2019). İki maddenin bağımlı olduğundan şüphelenmek için Linacre (2022), maddelerin artıkları arasındaki korelasyon katsayısının .7'den büyük olması gerektiğini belirtmektedir. Bunlardan hareketle, ölçeğin bu iki varsayımı ihlal edip etmediğini belirlemek için PCAR analizi gerçekleştirilmiş ve artıklar arasındaki korelasyon katsayıları hesaplanmıştır.

3.4.7. Madde Yanlılığı

Ölçek içerisinde yer alan maddelerin, örneklem farklı alt kümelerinde benzer şekilde işleyip işlemediği madde işleyiş farklılaşması analizi (differential item functioning [MİFA]) ile sınanmaktadır (Linacre, 2022). Ölçek maddelerinin değişmezlik sergilemesi ölçeğin yapı geçerliğini etkileyen önemli unsurlardan biridir. Bu değişmezlik ilkesi, maddelerin onaylanabilirlik açısından sıralamasının örneklem alt kümelerinde farklılaşmaması anlamına gelir (Boone vd., 2014). Ölçek maddelerinin değişmezliğini test etmek için örneklem içerisinde üniversite değişkeni açısından en kalabalık iki alt küme belirlenmiş ve Winsteps yazılımı ile MİFA gerçekleştirilmiştir. Örneklem iki alt kümesinde bir maddenin yanlılık gösterdiğine karar kılmak için maddenin iki gruptaki ölçüleri farkının mutlak değerinin .64'ten büyük ve elde edilen farkın ortaya çıkma olasılığının .05'ten küçük olması kıstas olarak önerilmektedir (Boone vd., 2014).

4. Bulgular

Bu bölümde esas uygulamadan elde edilen verilerin analizi sonucunda elde edilen bulgular, araştırma problemleri paralelinde oluşturulmuş başlıklar altında sunulmuştur.

4.1. Kategori Yapısının Ölçmedeki İşlevselliğine İlişkin Bulgular

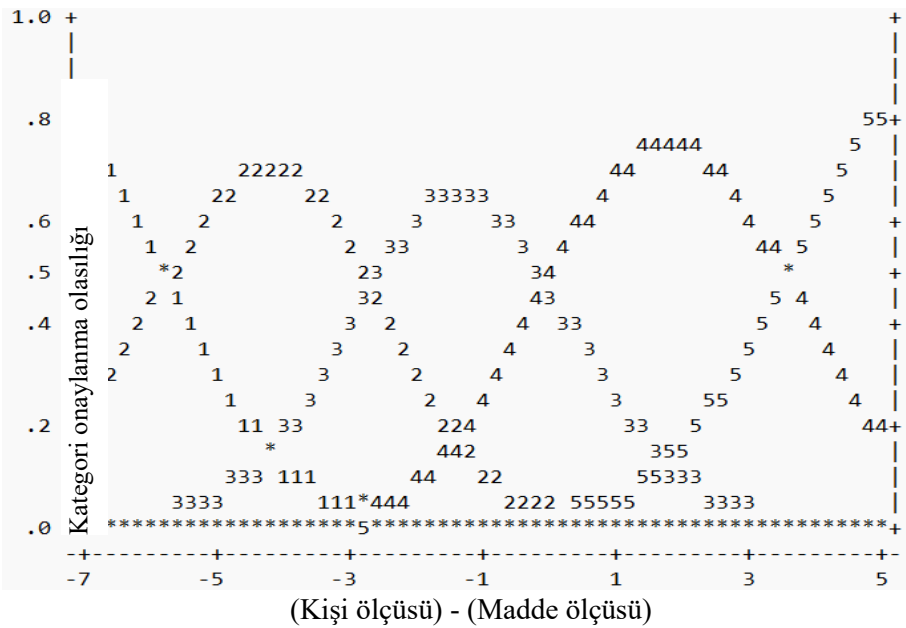
Farklı kategorileri onaylayan bireylerin ölçüleri ortalamasına ilişkin yapılan incelemede beklentinin aksini gösterecek bir bulgu ile karşılaşılmamıştır. Yalnızca M4 ve M7 kodlu maddelerde 1 ile 2 puanlık kategoriler için bir uyumsuzluk görülmüştür. Lakin ilgili maddelerin her ikisinde de 1 puanlık kategoriyi onaylayan sadece 1'er kişi, 2 puanlık kategoriyi onaylayan ise sırasıyla 40 ile 17 kişi bulunmaktadır. Dolayısıyla her iki madde için de ortaya çıkan uyumsuzluk 1 kişinin ölçüsü ile sırasıyla 40 ile 17 kişinin ölçülerinin mukayesesinden ileri gelmektedir.

Veri analizi kısmında belirtilen ve Linacre'nin (2002) önerdiği ilk 5 kıstasa yönelik elde edilen bulgular [Tablo 2'](#)de sunulmuştur. Elde edilen bulgular ilk beş kıstasın tümünün karşılandığını göstermektedir.

Tablo 2. İlk Beş Kıstasa Yönelik Elde Edilen Bulgular

Kategoriler (Puan)	Frekans (%)	Ortalama ölçü	Dış uyum (MnSq)	Kategori eşik (Andrich)
Çok az yeterliyim (1)	80 (1)	-4.04	1.51	—
Az yeterliyim (2)	671 (8)	-2.67	1.09	-5.75
Orta düzeyde yeterliyim (3)	2394 (29)	-0.38	0.91	-2.68
Yeterliyim (4)	4184 (50)	1.4	0.95	-0.03
Çok yeterliyim (5)	1051 (13)	2.91	1.04	3.54

Son kıstası değerlendirmek için kategori olasılık eğrileri oluşturulmuştur. Ortaya çıkan durum [Şekil 3](#)'te verilmiştir.

Şekil 3. Kategori Olasılık Eğrileri

[Şekil 3](#)'te gösterilen koordinat sisteminde düşey eksen olasılık değerini, yatay eksen ise kişi ölçüsü ile madde ölçüsü arasındaki farkı temsil etmektedir. Bu koordinat sistemi, bir katılımcının herhangi bir maddede bir kategoriye onaylama olasılığını katılımcının ölçüsü ile maddenin ölçüsü arasındaki farka dayalı olarak göstermektedir. [Şekil 3](#)'te görüldüğü üzere her bir kategorinin seçilme olasılığının en büyük olduğu bir aralık bulunmaktadır. Linacre'nin ifadesiyle tüm kategori eğrileri zirve oluşturmuştur.

4.2. Maddelerinin Dağılımına ve Ölçeğin Hedefleme Yeterliliğine İlişkin Bulgular

Gerçekleştirilen ölçüm sonundan ortaya çıkan kişi-madde haritası [Şekil 4](#)'te sunulmuştur.

Şekil 4. Kişi-Madde Haritası



Şeklin ortasındaki düşey çizgi kişilerin ve maddelerin ölçülerinin gösterildiği eş aralıklı ölçeği temsil etmektedir. Şeklin sol tarafında örneklemedeki bireylerin ölçülerinin dağılımı bulunmaktadır. Buradaki “#” sembolü 4 kişiyi, “.” sembolü ise 1, 2 veya 3 kişiyi temsil etmektedir. Şeklin solunda ve sağında bulunan M, S ve T harfleri ise sırasıyla ortalama, 1 standart sapma ve 2 standart sapma değerlerini göstermektedir. Kişiler kısmında, aşağıdan yukarıya doğru çıkıldıkça katılımcıların öz yeterlik ölçüleri artar. Diğer bir deyişle, yukarı doğru çıkıldıkça elde edilen ham puan yükselir. Sağ tarafta ise aşağıdan yukarıya doğru çıkıldıkça maddelerin onaylanması zorlaşır. Bir başka ifadeyle, maddenin elde ettiği toplam ham puan azalır.

Çalışmada ortaya çıkan kişi-madde haritası dikkate alındığında kişilerin ve maddelerin karşılıklı olarak konumlandığı değerlendirilmektedir. Gerçekleştirilen bir Rasch analizinde maddelerin ölçülerinin ortalaması varsayılan olarak 0 logit değerine kalibre edilmektedir. Bu çalışmada yapılan ölçümde kişi ölçülerinin ortalaması 0.7 logit olarak elde edilmiştir.

4.3. Madde Uyum İstatistiklerine İlişkin Bulgular

Gerçekleştirilen ölçüm sonucunda hesaplanan uyum istatistikleri ve nokta-ölçü korelasyon değerleri **Tablo 3**'te sunulmuştur. **Tablo 3**'te görüldüğü üzere maddelerin dış uyum istatistikleri 0.61–1.3, iç uyum istatistikleri 0.63–1.3, nokta-ölçü korelasyonları ise .48–.76 aralığında değişim göstermektedir.

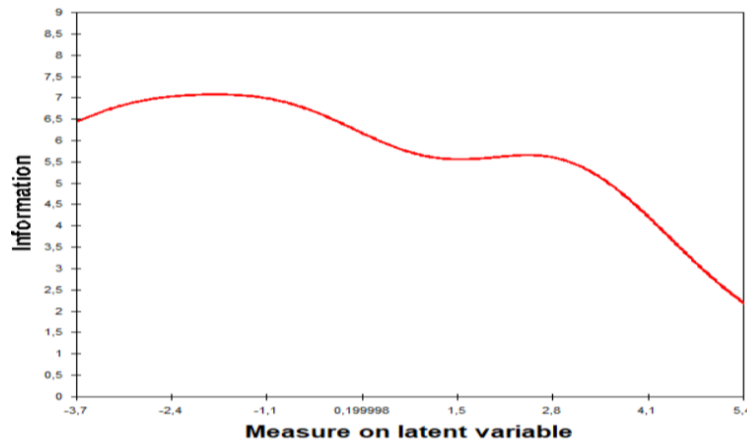
Tablo 3. Maddelerin Ölçüleri, Uyum İstatistikleri ve Nokta-Ölçü Korelasyon Değerleri

Madde	Madde Ölçüsü (SH)	Dış uyum (MnSq)	İç uyum (MnSq)	Nök	Madde	Madde Ölçüsü (SH)	Dış Uyum (MnSq)	İç uyum (MnSq)	Nök
M1	3.83 (0.08)	0.61	0.63	.76	M11	-0.50 (0.09)	1.03	1.05	.65
M2	3.32 (0.08)	0.64	0.67	.73	M12	-0.64 (0.09)	1.05	1.06	.63
M3	1.22 (0.08)	0.82	0.82	.74	M13	-0.72 (0.09)	1.23	1.26	.63
M4	0.72 (0.09)	1.26	1.23	.60	M14	-0.76 (0.09)	0.86	0.88	.66
M5	0.28 (0.09)	1.06	1.06	.60	M15	-0.82 (0.09)	1.23	1.25	.61
M6	0.07 (0.09)	1.29	1.30	.62	M16	-0.83 (0.09)	1.17	1.20	.54
M7	-0.10 (0.09)	0.77	0.79	.70	M17	-0.91 (0.09)	0.93	0.96	.62
M8	-0.14 (0.09)	0.75	0.76	.70	M18	-0.73 (0.09)	0.93	0.85	.48
M9	-0.36 (0.09)	0.91	0.92	.67	M19	-0.99 (0.09)	1.3	1.19	.49
M10	-0.45 (0.09)	1.08	1.11	.61	M20	-1.49 (0.09)	1.13	1.11	.61

Nök: Nokta-ölçü korelasyonu; SH: Standart hata

4.4. Güvenirlik ve Ayırt Ediciliğe İlişkin Elde edilen Bulgular

Çalışmada gerçekleştirilen ölçüm sonucunda ham puanların güvenirlik katsayısı (Cronbach's alfa) .92, kişilerin logit ölçülerinin güvenirlik katsayısı .91 ve maddelerin logit ölçülerinin güvenirlik katsayısı .99 olarak hesaplanmıştır. Elde edilen katsayılar alanyazında alt eşik olarak önerilen .7'den büyüktür (Boone vd., 2014). Gerçekleştirilen ölçümde kişilerin logit ölçüleri - 3.68 ile 5.38 arasında dağılmakta olup ölçüsü 4 logitten büyük olan 6 kişi bulunmaktadır. Ölçüleri farklı olan bireyleri testin ne hassasiyette ölçtüğünü belirlemek için bu aralıkta test bilgi fonksiyonunun grafiği oluşturulmuştur. Test bilgi fonksiyonu, ölçmenin standart hatasının karesinin çarpma işlemine göre tersi olarak tanımlandığından, standart hatanın 0.5 olduğu bir ölçüm noktasında bilgi fonksiyonu 4 değerine ulaşır (Boone ve Staver, 2020). [Şekil 5](#)'te görüldüğü üzere test örneklemdaki bireylerin hemen hemen tamamını standart hata 0.5'den küçük olacak şekilde ölçmüştür.

Şekil 5. Test Bilgi Fonksiyonu

Bununla birlikte madde ayırt edicilik indeksi 14.34, kişi ayırt edicilik indeksi 3.19 olarak hesaplanmıştır. Kişi ayırt edicilik indeksi ile hesaplanan tabaka değeri ise yaklaşık olarak 4'tür.

4.5. Tek Boyutluluk ve Yerel Bağımsızlık Kriterlerine Yönelik Bulgular

Bu çalışmada gerçekleştirilen PCAR analizi sonucunda örtük değişkendeki değişimden kaynaklı varyans toplam varyansın yaklaşık %57.1'ini açıklamakta olup artıkların korelasyon matrisinden hesaplanan birinci temel bileşenin özdeğeri 1.91 olarak hesaplanmıştır. Diğer taraftan, yerel bağımsızlık varsayımı için madde artıkları arasındaki korelasyon katsayıları hesaplanmış ve $-.25$ ile $.41$ arasında değiştiği saptanmıştır.

4.6. Madde Yanlılığına İlişkin Bulgular

Örneklem içerisinde üniversite değişkeni açısından en kalabalık iki alt küme Trabzon ve 9 Eylül üniversiteleridir. Maddelerin yanlılık gösterip göstermediğini belirlemek amacıyla bu İki üniversite dikkate alınarak her bir madde için “Madde ölçüsü iki grupta eşittir” hipotezi test edilmiştir. MİFA sonucunda elde edilen bulgular [Tablo 4](#)'te sunulmuştur. [Tablo 4](#)'te görüldüğü üzere tüm maddelerin iki grupta belirlenen ölçüleri arasındaki farkın mutlak değeri 0.64'ten küçük olup tümü için elde edilen olasılık değeri anlamlılık sınırından ($.05$) büyüktür.

Tablo 4. Trabzon ve 9 Eylül Üniversiteleri Alt Grupları için MİFA Bulguları

Madde	Trabzon		9 Eylül		MİF Fark	<i>t</i>	<i>p</i>
	MİF ölçüsü	SH	MİF ölçüsü	SH			
M1	3.73	0.20	3.69	0.22	0.04	0.14	.89
M2	3.47	0.19	3.18	0.21	0.29	0.99	.33
M3	1.31	0.19	1.22	0.22	0.09	0.31	.76
M4	0.77	0.20	1.19	0.20	0.42	-1.42	.16
M5	0.22	0.20	0.28	0.23	0.06	-0.21	.83
M6	0.18	0.20	-0.12	0.23	0.30	0.97	.33
M7	0.22	0.20	0.09	0.23	0.13	0.42	.68
M8	-0.41	0.22	0.09	0.23	0.50	-1.61	.11
M9	-0.03	0.20	-0.56	0.24	0.53	1.7	.09
M10	-0.62	0.21	-0.28	0.23	0.34	-1.08	.28
M11	-0.75	0.21	-0.28	0.33	0.47	-1.49	.14
M12	-0.71	0.21	-0.96	0.24	0.25	0.79	.43
M13	-1.02	0.21	-0.78	0.24	0.32	-0.72	.47
M14	-0.79	0.21	-0.67	0.21	0.12	-0.39	.70
M15	-1.02	0.20	-0.90	0.24	0.12	-0.36	.72
M16	-0.53	0.21	-0.67	0.24	0.14	0.44	.66
M17	-0.84	0.21	-1.19	0.24	0.35	1.10	.27
M18	-0.73	0.21	-0.90	0.24	0.17	0.53	.60
M19	-0.99	0.21	-1.13	0.24	0.14	0.45	.66
M20	-1.52	0.22	-1.37	0.25	0.15	-0.44	.66

4. Sonuç, Tartışma ve Öneriler

Araştırmada geçerlik açısından incelenen durumlardan biri ölçek maddelerine tepki vermede benimsen kategori yapısının işlevselliği idi. Elde edilen bulgular kategorilerin onaylanma frekansına, ortalama ölçülerine, uyum indeksine, kategori eşik değerlerine ve kategorilerin olasılık eğrilerine ilişkin ileri sürülmüş kıstasların karşılandığını ortaya koymuştur. Sonuç olarak benimsenen kategori yapısı ölçme açısından iyi işlev gösterdiğinden ölçeğin gelecekteki araştırmalarda mevcut kategori yapısı ile kullanılması önerilmektedir.

Çalışmada ele alınan bir diğer husus gerçekleştirilen ölçümün güvenilirlik ve ayırt edicilik açısından incelenmesiydi. Kişilerin ham puanları üzerinde hesaplanan güvenilirlik katsayısı gerçekleştirilen ölçümün yüksek seviyede iç tutarlılığa sahip olduğuna işaret etmektedir. Hem kişilerin hem de maddelerin logit ölçüleri üzerinden hesaplanan güvenilirlik katsayıları alanyazında ileri sürülen alt eşikten yüksek çıkmıştır. Bu durum, ölçek gelecekte farklı bir örnekleme uygulanacak olsa madde ölçülerinin büyük ihtimalle burada elde edilen ölçülere yakın olacağı anlamına gelmektedir (Bond vd., 2021). Diğer taraftan madde ayırt edicilik değeri alanyazında önerilen eşik üzerinde hesaplanmıştır. Buna bağlı olarak maddelere ilişkin tabaka değeri de oldukça yüksek elde edilmiştir. Bu değerler, ölçek maddelerinin bir bütün olarak matematik dilini öğretmeye ilişkin öz yeterlik değişkeninin geniş bir aralığını ölçebileceğine işaret etmektedir. Bununla birlikte kişi ayırt edicilik indeksi temelinde hesaplanan kişilere yönelik tabaka değeri yaklaşık olarak 4 bulunmuştur. Hesaplanan bu değer, ölçeğin kişileri ilgili niteliğe sahiplik açısından 4 seviyeye ayırabileceğini söylemektedir.

Ölçek maddelerinin kişi-madde haritasındaki dağılımı dikkate alındığında maddelerin ölçülen değişkenin geniş bir aralığında dağıldığı görülmektedir. Bununla birlikte, ölçeğin onaylanması en zor ikinci maddesi ile 3. maddesi arasında kayda değer bir boşluk bulunmaktadır. Bu durum, ölçüsü bu bölgede yer alan bireyleri birbirinden ayırt etmede ölçeğin hassaslık gösteremeyeceği anlamına gelmektedir. Bu sebepten ötürü ölçeğin ölçüsü bu bölgeye düşecek madde veya maddeler ile zenginleştirilmesi ileriki araştırmalarda ele alınması gereken bir ihtiyaçtır. Alanyazındaki öneriler temelinde örnekleme hedefleme yeterliği açısından ölçeğin orta düzeyde bir performans sergilediği değerlendirilmektedir. Gerçekleştirilen ölçümde kişilerin ölçüleri ortalaması, maddelerin ölçüleri ortalamasından 0.7 logit kadar fazla hesaplandığından ölçek örnekleme hedefleme açısından orta düzeyde isabetsizlik sergilemiştir. Dolayısıyla, ölçek maddelerinin bir bütün olarak onaylanabilirliği örneklem için bir nebze kolaydır. Daha önce ifade edilen zenginleştirmenin gerçekleştirilmesi ölçeğin hedefleme yeterliğine de olumlu katkı sunacaktır.

Uyum açısından ele alındığında, ölçek maddelerinin her iki uyum istatistikleri de önerilen aralık içerisinde yer almaktadır. Dolayısıyla ölçek maddelerine verilen yanıtlar modelin beklentisinden anlamlı düzeyde sapma sergilememiştir. Bu durum, ölçekte yer alan maddelerin bütünsel olarak odaklanılan niteliğin ölçümüne hizmet ettiğinin bir göstergesidir (Oon ve Subramaniam, 2011). Bunun yanı sıra, maddeler için hesaplanan nokta-ölçü korelasyonları da yüksek pozitif değerlerden oluşmaktadır. Elde edilen değerler, ölçeğin her bir maddesinin eş yönlü davrandığını dolayısıyla ölçülmesi amaçlanan nitelik ile alakalı olduğunu söylemektedir.

RÖT'ün iki temel varsayımı olan tek boyutluluk ve yerel bağımsızlık için sırasıyla PCAR analizi gerçekleştirilmiş ve artıkların korelasyon matrisi incelenmiştir. PCAR analizi sonucunda Rasch ölçülerinden kaynaklı varyansın toplam varyanstaki oranı ve artıkların korelasyon matrisinden hesaplanan birinci temel bileşenin özdeğeri ölçeğin Rasch boyutu haricinde başka bir boyut içermediğine işaret etmektedir. Dolayısıyla hazırlanan ölçek, tek bir örtük değişkenin ölçümüne hizmet etmektedir. Bununla birlikte artıklar arasındaki korelasyon katsayısının önerilen üst eşik değerinden küçük olmasından ötürü ölçeğin yerel bağımsızlık kriterini ihlal etmediği değerlendirilmektedir.

Maddelerin yanlılık gösterip göstermediğini sınamak için üniversite değişkeni açısından örneklem içerisinde hacimce en büyük iki grup alınarak MİFA gerçekleştirilmiştir. Her iki gruptan hesaplanan madde ölçüleri arasındaki fark alanyazında önerilen eşik değerden düşük bulunmuştur. Bu durum ölçek maddelerin farklı örneklemlerde benzer davrandığını, bir başka ifadeyle kişi-madde haritasında yer alan konumlarının anlamlı bir farklılık oluşturmadığı anlamına gelmektedir.

Çalışmadan elde edilen bulgular ve alanyazında önerilen kıstaslar göz önüne alındığında, bu çalışmada geliştirilen ölçeğin kabul edilebilir düzeyde geçerli ve güvenilir bir ölçme işlemi gerçekleştirdiğine kanaat getirilmiştir. Sonuç olarak, ilköğretim matematik öğretmeni adaylarının matematik dilini öğretmeye ilişkin öz yeterliklerini belirlemeyi amaç edinen çalışmalarda kullanılabilir.

Etik Kurul İzin Bilgisi

Bu araştırma, Trabzon Üniversitesi Sosyal ve Beşeri Bilimler Bilimsel Araştırma ve Yayın Etiği Kurulunun 24.04.2024 tarihli E-81614018-050.04-2400019593 sayılı kararı ile alınan izinle yürütülmüştür.

Yazar Çıkar Çatışması Bilgisi

Yazarların beyan edeceği bir çıkar çatışması yoktur.

Yazar Katkısı

Erdem Çekmez: Literatür taraması, veri toplama, araştırma raporunun yazılması, verinin analizi ve yorumlanması. **Mustafa Güler:** Literatür taraması, veri toplama, kavramsal çerçevenin oluşturulması, araştırma raporunun yazılması. **Taha Ertuğrul Kuzu:** Literatür taraması, kavramsal çerçevenin oluşturulması.

Orcid

Erdem Çekmez  <https://orcid.org/0000-0001-8684-2820>

Mustafa Güler  <https://orcid.org/0000-0002-4082-7585>

Taha Ertuğrul Kuzu  <https://orcid.org/0000-0001-7422-0231>

KAYNAKÇA

- Açıl, E. ve Zeybek, Z. (2017). Öğrencilerin matematiksel dili kullanma ve anlama becerisi ile öğretmenlerinin öğrencilerin matematiksel dili nasıl kullandıklarını fark edebilme yeteneği. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 42, 87-107.
- Adler, J. (1999). The dilemma of transparency: Seeing and seeing through talk in the mathematics classroom. *Journal for Research in Mathematics Education*, 30(1), 47-64. <https://doi.org/10.2307/749629>
- Akcan, F. (2023). *Sosyal bilgiler dersinde dijital güvenlik yeterliği için geliştirilen çevrim içi destekleyici öğrenme aracının öğrencilerin öz yeterlik algılarına etkisi* (Tez No: 776088) [Doktora tezi, Gazi Üniversitesi]. Yök Tez Merkezi.
- 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>
- Araujo, Z. d., Roberts, S. A., Willey, C., & Zahner, W. (2018). English learners in K–12 mathematics education: A review of the literature. *Review of Educational Research*, 88(6), 879-919. <https://doi.org/10.3102/0034654318798093>
- Baber, R. L. (2011). *The language of mathematics: Utilizing math in practice*. John Wiley & Sons.
- Bali-Çalikoğlu, G. (2002). Matematik öğretiminde dil ölçeği. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 23, 57-61.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191-215. <https://doi.org/10.1037/0033-295X.84.2.191>
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman and Company.

- Barton, B. (2008). *The language of mathematics: Telling mathematical tales*. Springer. <https://doi.org/10.1007/978-0-387-72859-9>
- Barwell, R. (2016). Formal and informal mathematical discourses: Bakhtin and Vygotsky, dialogue and dialectic. *Educational Studies in Mathematics*, 92(3), 331-345. <https://doi.org/10.1007/s10649-015-9641-z>
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Bond, T. G., Yan, Z., & Heene, M. (2021). *Applying the Rasch model: Fundamental measurement in the human sciences* (4th ed.). Routledge
- Boone, W. J. (2020). Rasch basics for the novice. In M. S. Khine (Ed.), *Rasch measurement: Applications in quantitative educational research* (pp. 9-30). Springer. <https://doi.org/10.1007/978-981-15-1800-3>
- Boone, W. J., & Staver, J. R. (2020). *Advances in Rasch analysis in the human sciences*. 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*. Springer. https://doi.org/10.1007/978-94-007-6857-4_4
- Burns, M. (2006). Marilyn Burns on the language of math. *Instructor*, 115(7), 41-43.
- Büyüköztürk, Ş., Kılıç-Çakmak, E., Akgün, Ö., Karadeniz, Ş. ve Demirel, F. (2019). *Bilimsel araştırma yöntemleri*. 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. *Psychological Assessment*, 23(2), 582-562. <https://doi.org/10.1037/a0022484>
- Çekmez, E., Tuti, G. ve Terzi, S. Y. (2023). İlkokul 3. ve 4. sınıf öğrencilerine yönelik matematik kaygı ölçeğinin psikometrik özelliklerinin incelenmesi. *Anadolu University Journal of Education Faculty*, 7(4), 819-841. <https://doi.org/10.34056/aujef.1254390>
- Çepni, S. (2007). *Araştırma ve proje çalışmalarına giriş*. Celepler Matbaacılık.
- Duncan, P. W., Bode, R. K., Lai, S. M., & Perera, S. (2003). Rasch analysis of a new stroke-specific outcomescale: 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)
- Durkin, K., & Shire, B. (1991). Lexical ambiguity in mathematical contexts. In K. Durkin & B. Shire (Eds.), *Language in mathematical education* (pp. 71-84). Open University Press.
- Erath, K., Prediger, S., Quasthoff, U. M., & Heller, V. (2018). Discourse competence as important part of academic language proficiency in mathematics classrooms: The case of explaining to learn and learning to explain. *Educational Studies in Mathematics*, 99(2), 161-179. <https://doi.org/10.1007/s10649-018-9830-7>
- 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: Fundamental techniques* (pp.83-102). Routledge. <https://doi.org/10.4324/9781315187815>
- Gökmen, A.D.Y. ve Ulusoy, M. (2022). Kelime bilgisi ve anlama becerilerinin anlatım becerilerindeki rolü: Farklı yol analizi denemeleri. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 56, 372-400. <https://doi.org/10.9779.pauefd.1026827>
- Gürefe, N. (2018). Mathematical language skills of mathematics prospective teachers. *Universal Journal of Educational Research*, 6(4), 661-671. <https://doi.org/10.13189/ujer.2018.060410>

- Gray, V. (2004). *The language of mathematics: A functional definition and development of an instrument to measure teacher perceived self-efficacy* [Doctoral dissertation, Oregon State University]. <http://hdl.handle.net/1957/31694>
- Halliday, M. A. K. (1978). *Language as social semiotic*. Edward Arnold.
- Kabael, T. ve Yayan, B. (2017). Effect of self-evaluation on pre-service mathematics teachers' self-efficacy in language of mathematics. *Anadolu Journal of Educational Sciences International*, 7(1), 1-34. <https://doi.org/10.18039/ajesi.292575>
- Kuzu, T. (2023). Multilingual meaning making – An explorative study of German-Turkish learners' translanguaging processes regarding the part-whole-concept. *Journal für Mathematik-Didaktik*, 44(4), 325-353. <https://doi.org/10.1007/s13138-023-00219-z>
- Kuzu, T., & Ratzke, C. (2024). Explanatory videos in the context of learning processes: An interdisciplinary interpretative interaction analysis of production and reflection processes. *Journal of Pedagogical Research*, 8(2), 227-259. <https://doi.org/10.33902/JPR.202426555>
- Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3(1), 85-106.
- Linacre, J. M. (2022). *A user guide to Winsteps Ministep Rasch model computer programs*. <https://www.winsteps.com/a/Winsteps-Manual.pdf>
- Millî Eğitim Bakanlığı. (2018a). *Matematik dersi öğretim programı (İlkokul ve ortaokul 1,2,3,4,5,6,7, ve 8. sınıflar)*. Millî Eğitim Bakanlığı.
- Millî Eğitim Bakanlığı. (2018b). *Ortaöğretim matematik dersi (9, 10, 11 ve 12. sınıflar) öğretim programı*. Millî Eğitim Bakanlığı.
- Millî Eğitim Bakanlığı. (2024a). *Ortaöğretim matematik dersi öğretim programı (Hazırlık, 9, 10, 11 ve 12. sınıflar)*. Millî Eğitim Bakanlığı.
- Millî Eğitim Bakanlığı. (2024b). *Ortaokul matematik dersi öğretim programı (5, 6, 7 ve 8. sınıflar)*. Millî Eğitim Bakanlığı.
- Moleko, M. M. (2021). Teachers' perspectives on addressing linguistic factors affecting visualisation of mathematics word problems. *Eurasia Journal of Mathematics, Science and Technology Education*, 17(11), em2029. <https://doi.org/10.29333/ejmste/11248>
- Moore, R. C. (1994). Making the transition to formal proof. *Educational Studies in Mathematics*, 27(3), 249-266. <https://doi.org/10.1007/bf01273731>
- Morgan, C. (2020). Mathematical language. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 540-543). Springer. <https://doi.org/10.1007/978-3-030-15789-0>
- Moschkovich, J. (2015). Academic literacy in mathematics for english learners. *The Journal of Mathematical Behavior*, 40, 43-62. <https://doi.org/10.1016/j.jmathb.2015.01.005>
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. NCTM.
- 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. <https://doi.org/10.1080/09500693.2010.511297>
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.
- 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 (Ed.), *Applications of Rasch measurement in learning environments research* (pp. 119-139). Sense Pub.
- Özgen, K. ve Bindak, R. (2008). Matematik okuryazarlığı öz-yeterlik ölçeğinin geliştirilmesi. *Kastamonu Eğitim Dergisi*, 16(2), 517-528.

- Peng, P., Lin, X., Ünal, Z., Lee, K., Namkung, J., Chow, J., ... & Sales, A. (2020). Examining the mutual relations between language and mathematics: A meta-analysis. *Psychological Bulletin*, 146(7), 595-634. <https://doi.org/10.1037/bul0000231>
- Pimm, D. (1987). *Speaking Mathematically: Communication in mathematics classrooms*. Routledge.
- Prediger, S. (2019). Investigating and promoting teachers' expertise for language-responsive mathematics teaching. *Mathematics Education Research Journal*, 31, 367-392. <https://doi.org/10.1007/s13394-019-00258-1>
- Prediger, S., Kuzu, T., Schüler-Meyer, A., & Wagner, J. (2019). One mind, two languages – separate conceptualisations? A case study of students' bilingual modes for dealing with language-related conceptualisations of fractions. *Research in Mathematics Education*, 21(2), 188-207. <https://doi.org/10.1080/14794802.2019.1602561>
- Riccomini, P. J., Smith, G. W., Hughes, E. M., & Fries, K. M. (2015). The language of mathematics: The importance of teaching and learning mathematical vocabulary. *Reading & Writing Quarterly*, 31(3), 235-252. <https://doi.org/10.1080/10573569.2015.1030995>
- Schleppegrell, M. J. (2007). The linguistic challenges of mathematics teaching and learning: A research review. *Reading & Writing Quarterly*, 23(2), 139-159. <https://doi.org/10.1080/10573560601158461>
- Seçer, İ. (2018). *Psikolojik test geliştirme ve uyarlama süreci: SPSS ve LISREL uygulamaları* (2. Baskı). Anı Yayıncılık.
- Smith Jr, E. V., Conrad, K. M., Chang, K., & 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., & Muñiz, J. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema*, 30(2), 149-158. <https://doi.org/10.7334/psicothema2018.33>
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson.
- Tennant, A., & Conaghan, P. (2007). The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Care and Research*, 5(8), 1358–1362. <https://doi.org/10.1002/art.23108>
- Uptegrove, E. B. (2015). Shared communication in building mathematical ideas: A longitudinal study. *Journal of Mathematical Behavior*, 40, 106–130. <https://doi.org/10.1016/j.jmathb.2015.02.001>
- Wilkinson, L. C. (2018). Teaching the language of mathematics: What the research tells us teachers need to know and do. *The Journal of Mathematical Behavior*, 51, 167-174. <https://doi.org/10.1016/j.jmathb.2018.05.001>
- Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8(3), 370.
- Wright, B.D., & Stone, M. H. (2004). *Making measures*. Phaneron Press
- 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>

Ek A. Ölçek Maddeleri

M5. Hem günlük konuşma dilinde hem de matematik dilinde kullanılan kelimelerin (bu türden kelimelere örnek: benzer, rasyonel, yamuk) anlamları arasındaki ayrımı öğrencilere fark ettirecek açıklamalar yapabilirim.

M18. Öğrencilerin matematiksel terimleri uygun biçimde kullanıp kullanamadıklarını fark edebilirim.

M14. Matematiksel ifadeleri kurallara uygun biçimde yazma bakımından öğrencilerime örnek olabilirim.

M16. Öğrencilerimin defterlerine yazacakları matematiksel ifadelerin kurallara uygun olmasını sağlayabilirim.

M20. Denklem oluştururken harflerin veya simgelerin ne amaçla kullanıldığını öğrencilerime kavratabilirim.

M9. Öğrencilere matematiksel terimleri uygun biçimde kullanma alışkanlığını kazandırabilirim.

M11. Öğrencilerin matematik dilinde yazılmış metinleri sözel olarak ifade etme yeterliklerini geliştirebilirim.

M15. Grafiklerin matematikteki önemini öğrencilerime fark ettirebilirim.

M4. Öğrencilerimin matematiksel düşüncelerini sınıf arkadaşları ile tartışmaya istekli olmalarını sağlayabilirim.

M8. Matematiksel kavramların farklı temsilleri arasındaki ilişkiyi öğrencilerime kavratabilirim.

M10. Sınıf ortamını öğrencilerin matematiksel düşüncelerini sözlü olarak paylaşmalarını teşvik edecek biçimde düzenleyebilirim.

M17. Öğrencilerin matematik dilini kullanırken yaptıkları hataları fark edebilirim.

M7. Öğrencilerin matematiksel bilgiyi yazılı olarak ifade etme yeterliklerini geliştirebilirim.

M13. Öğrencilere grafik okuma becerisini kazandırabilirim.

M6. Matematiksel bilgiyi açıklarken kelimeleri öğrencilerin yaş düzeyine uygun olacak şekilde seçebilirim.

M3. Öğrencilerin sözel problemlerde verilen ilişkileri matematiksel olarak ifade etme becerilerini geliştirecek etkinlikler tasarlayabilirim.

M2. Öğrencilerin matematik dili becerilerini geliştirme amacıyla tasarlanmış etkinlikler bulmak istersem nerede arayacağımı biliyorum.

M19. Geometrik şekilleri, şekli oluşturan elemanların ölçülerini dikkate alarak çizme hususunda öğrencilerime örnek olabilirim (örneğin: bir açının ölçüsü 90 dereceden büyük ise açığı geniş açı gibi görünecek şekilde çizmek).

M12. Öğrencilerimin matematiksel sembolleri ve notasyonları doğru okumalarına özen gösteririm.

M1. Öğrencilerin matematik dili yeterliklerini ölçme amacıyla tasarlanmış değerlendirme türünden etkinliklere ilişkin bilgi sahibiyim.

Development of the Mathematical Language Teaching Self-Efficacy Scale for Pre-service Elementary Mathematics Teachers

Erdem ÇEKMEZ^{1*}, Mustafa GÜLER¹, Taha Ertuğrul KUZU²

¹ Trabzon University, Fatih Faculty of Education, Department of Sci. and Math. Education, Türkiye

² University of Education, Schwäbisch Gmünd, Germany

Abstract: *This study aimed to develop a scale to determine the self-efficacy of pre-service primary mathematics teachers regarding teaching the language of mathematics. The participants of the research consisted of 419 teacher candidates studying in the final semester of the relevant program at 14 universities. The extent to which the scale provides valid and reliable measurement was examined in terms of item fit statistics, the functionality of the adopted category structure, reliability and discrimination indices for persons and items, sample targeting adequacy, the hierarchy formed by the items, item bias status, unidimensionality, and local independence components. Data analysis was conducted using Winsteps® software based on the rating scale model within the framework of Rasch measurement theory. The findings revealed that the category structure was functional, item fit statistics were within the ranges suggested in the literature, the criteria for unidimensionality and local independence were met, the reliability and discrimination indices for persons and items were appropriate, and the items did not exhibit bias. However, it was determined that the scale demonstrated a moderate level of mistargeting concerning the sample. Consequently, the scale provides reasonably valid and reliable measurement and can be utilized in future research.*

Article Details

Research Article

Received

27/06/2024

Accepted

19/03/2025

Keywords

Mathematical
language,
Self-efficacy,
Prospective
mathematics
teachers.

1. Introduction

Language plays a significant role in the acquisition of mathematical knowledge. It serves as a medium for teachers to explain mathematical concepts and methods to students, to interpret mathematical texts, and to facilitate thinking during mathematical activities such as problem-solving (Gökmen & Ulusoy, 2022). Moreover, language enables students to engage in the exchange of ideas with both their peers and their teacher. Acknowledging the importance of language-mediated communication in mathematics learning, the middle school mathematics curriculum in Turkey emphasizes the need to provide students with opportunities to solve problems by effectively utilizing language and symbols, articulate their thoughts verbally, and communicate via mathematical representations (Ministry of National Education [MoNE], 2024a).

* *Sorumlu Yazar:* Erdem ÇEKMEZ *E-mail:* erdemcekmez@trabzon.edu.tr *Address:* Trabzon University, Fatih Faculty of Education, 61300, Trabzon, Türkiye

The copyright of the published article belongs to its author under CC BY 4.0 license. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>

In the teaching process that takes place in mathematics classrooms, language emerges in two distinct forms: the everyday language that facilitates general communication and the mathematical language that mediates the representation and communication of mathematical knowledge. Mathematical language is a form of communication that enables mathematicians and students to express mathematical ideas clearly (Moore, 1994). The use of mathematical language is essential for understanding mathematical concepts and interacting with them (Açıl & Zeybek, 2017; Erath *et al.*, 2018; Prediger *et al.*, 2019). Furthermore, it allows individuals to describe and analyze mathematical relationships and to concretize their mathematical reasoning during problem-solving processes (Barwell, 2016).

For students to achieve meaningful learning regarding the mathematical content addressed, their proficiency in mathematical language must continuously develop throughout the teaching process. One aspect of this expected development is the expansion of students' domain-specific vocabulary and terminology in proportion to their educational level. Mathematical vocabulary plays a key role in both students' understanding of mathematical ideas and their effective use of these ideas in communication (Riccommini *et al.*, 2015). In line with this perspective, Uptegrove (2015) has demonstrated that students who meaningfully comprehend the symbolic words specific to mathematical language can develop a deep and robust understanding of mathematical knowledge. However, mathematical language is not merely about learning mathematical words and notations. It also encompasses various actions, such as imitating the style used to express mathematical knowledge, representing mathematical knowledge across different representational domains, and concretizing the thoughts occurring in one's mind on paper to facilitate the process of mathematical reasoning (Moschkovich, 2015; Peng *et al.*, 2020). For this reason, in addition to its communicative role, mathematical language also has a cognitive function, considering the fact that language mediates both learning and thinking (Erath *et al.*, 2018; Kuzu, 2023; Kuzu & Ratzke, 2024).

The effective teaching of mathematics largely depends on the correct use of mathematical language, which encompasses rules, concepts, symbols, and terms (Gürefe, 2018). Unlike everyday spoken language, this language involves the use of words with clearly defined meanings, its own syntactic rules, and a distinctive style of expression. For instance, the definitions of mathematical concepts include technical terms, dense noun phrases, conjunctions with specific meanings, and implicit logical relationships (Schleppegrell, 2007). With all these aspects, mathematical language poses challenges for students in understanding mathematical knowledge and solving problems. Some of these challenges include understanding words specific to mathematics that are not found in everyday language, distinguishing between words that exist in both everyday and mathematical language but have different meanings, and comprehending mathematics-specific noun phrases (Wilkinson, 2018). It has been noted that students who fail to overcome these challenges tend to explain mathematical knowledge in ways that are meaningful only to themselves (Moschkovich, 2015). In such cases, the lack of shared meaning makes it difficult for teachers to provide appropriate feedback. For these reasons, teachers need to be careful and clear when using mathematical language to support students' learning (Adler, 1999; Araujo *et al.*, 2018). Additionally, it is recommended that teachers rephrase and correct any inaccurate expressions in students' verbal explanations (Wilkinson, 2018). Despite these recommendations, Moleko (2021) found that teachers often fail to pay attention to linguistic factors during instruction related to problem-solving. This oversight makes it more difficult for students to effectively visualize and solve problems.

Although it is emphasized that mathematics teachers should design their teaching processes in ways that support students' proficiency in mathematical language, it is noted that this often does not occur (Baber, 2011; Durkin & Shire, 1991; Prediger, 2019). Gray (2004) argues that

teachers' self-efficacy in teaching this skill plays a significant role in whether or not they incorporate the teaching of mathematical language into their classroom practices. In this context, Gray suggests that if a teacher has low confidence in their ability to teach mathematical language, it is highly likely that they will refrain from making efforts in this direction. Based on this, the attitudes that prospective mathematics teachers will adopt toward teaching mathematical language in their future instructional practices are closely tied to their self-efficacy in this area. To assess the self-efficacy of prospective mathematics teachers regarding this issue, there is a need for measurement tools that are both valid and reliable at an acceptable level.

One of the measurement tools that can be evaluated in the field of mathematical language was developed by Kabael and Yayan (2017). However, this tool aims to measure prospective mathematics teachers' self-efficacy in using mathematical language rather than their self-efficacy in teaching it. For instance, one of the items in the scale reads, "I can express a proposition in mathematical language." Therefore, the scale developed in this study differs from Kabael and Yayan's scale in this respect.

Another existing scale was developed by Özgen and Bindak (2008) to determine prospective mathematics teachers' self-efficacy in mathematical literacy. The items in this scale focus on evaluating the broader concept of mathematical literacy rather than the teaching of mathematical language. Similarly, Bali-Çalikoğlu (2002) developed a scale to determine prospective mathematics teachers' views on the role of language in mathematics teaching. However, upon examining the items in this scale, it was concluded that they are not directly related to the latent variable targeted for measurement in this study.

A review of the literature revealed that no existing measurement tool specifically addresses the latent variable targeted in this study. Developing such a tool would not only help determine prospective teachers' self-efficacy but also, depending on the endorsability levels of the scale items, uncover the sub-competencies in which prospective teachers feel inadequate. This, in turn, could provide insights into the type of support that should be offered to prospective teachers during their pre-service training. Based on this rationale and significance, the aim of this study is to develop a scale to measure prospective teachers' self-efficacy in teaching mathematical language and to examine its psychometric properties within the framework of the Rasch Measurement Theory (RMT).

2. Conceptual Framework

2.1. Mathematical Language

Mathematical language is a form of communication that facilitates the concretization and sharing of mathematical knowledge with the aid of everyday spoken language. Although it is referred to as a "language," mathematical language differs from the languages people use for daily communication in several ways. For instance, no one has mathematics as their native language. Additionally, while individuals can speak their native language fluently without being literate, fluency in mathematical language requires mathematical literacy as a prerequisite. Furthermore, mathematical language relies on a native language to fulfill its communicative function.

Burns (2006) highlights that learning mathematical language is similar to learning a foreign language. However, he also emphasizes a key distinction: to learn mathematical language, one must first understand the concepts that will be discussed. This underscores the interdependence between conceptual understanding and language acquisition in mathematics. In addition, the most distinctive characteristic that sets mathematical language apart from everyday spoken language is its precision and freedom from ambiguity. These characteristics

make mathematical language a powerful tool for conducting flawless reasoning and defining technical terms (Baber, 2011).

Although the term "mathematical language" is frequently used in the mathematics education literature, there is no universally agreed-upon definition of the concept. For instance, Barton's (2008) book titled *The Language of Mathematics* does not provide a clear definition of mathematical language. Similarly, NCTM (2000) refers to the effective use of mathematical language within the framework of mathematical literacy, it does not offer an explicit definition.

In Turkey, the 2018 primary and middle school mathematics curriculum explicitly includes the term "mathematical language," yet it does not provide a definition (MoNE, 2018a). On the other hand, the high school mathematics curriculum does not mention the term at all (MoNE, 2018b). In the 2024 middle school curriculum (MoNE, 2024b), the term "mathematical language" is absent, whereas the high school curriculum (MoNE, 2024a) includes the term "mathematical language" but, like the 2018 curriculum, does not provide a definition. This suggests an implicit assumption that everyone understands the term "mathematical language" in the same way, despite the lack of a clear and shared definition.

When examining the mathematics education literature, it is evident that various researchers have attempted to explain or define mathematical language. In this context, Baber (2011) describes the language of mathematics as the notational forms used in mathematical practice and states that texts written in mathematical language consist of four fundamental elements: values, variables, functions, and expressions. Halliday (1978), on the other hand, approaches mathematical language as a style and defines it as the use of everyday spoken language for mathematical purposes, explicitly excluding symbolism from the language of mathematics. Morgan (2020) characterizes mathematical language by listing its distinctive features. These features include the use of domain-specific vocabulary, the ability to derive mathematical terms from verbs (e.g., deriving the term "translation"), the inclusion of algebraic expressions, and the utilization of different semiotic systems, such as the Cartesian coordinate system. Similarly, Pimm (1987) identifies the characteristics of mathematical language as including domain-specific meanings, the presence of entities symbolized by mathematical symbols, and its unique syntactic rules.

Gray (2004), in his doctoral study, emphasized the lack of a clear definition of mathematical language and aimed to provide a functional definition. As a result of his investigation, he defined mathematical language in the context of the English language as follows: "In English speaking classrooms, the language of mathematics is used in combination with English and a set of symbolic forms. This language refers to the specific words, symbols, graphs, and conventions used when discussing or writing about mathematics." (p. 21). Gray stated that he shared this proposed definition with experienced researchers in the field of mathematics education, who expressed their agreement that the definition was appropriate.

As a result of the literature review on the definition of mathematical language, it has been understood that mathematical language possesses its own unique notations, terms, and conventions; shares words with everyday spoken language; utilizes different representational and semiotic systems; has domain-specific syntactic rules; and that communication plays a significant role in its development. These characteristics were taken into account in the creation of the item pool for this study.

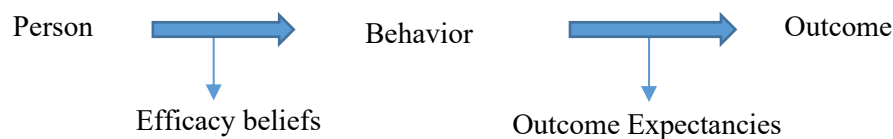
2.2. Self-Efficacy

Self-efficacy, first introduced by Albert Bandura, is a concept that asserts the behaviors exhibited are connected to individuals' beliefs and emotions (Akcan, 2023). Bandura (1997)

argues that human agency operates within an interconnected causal structure that involves triadic reciprocal causation. The triadic reciprocal causation structure consists of behavior, personal factors—whether cognitive, affective, or biological—and the environment. The prediction of this network of relationships is that individuals' choices are influenced not only by environmental factors but also by their personal evaluations of their ability to execute a chosen course of action.

Among the determinants of human agency, self-efficacy is defined by Bandura (1997, p. 21) as "a judgment of one's capability to organize and execute given types of performances." When an individual encounters a problem, their belief in their own self-efficacy becomes a decisive factor in whether they will attempt to solve the problem, how much effort they will exert in the process, and how much resilience they will demonstrate in overcoming potential difficulties (Bandura, 1977). Bandura (1997) further emphasizes that individuals' expectations regarding the outcomes of an action are largely dependent on their evaluations of how well they can perform that action.

Figure 1. *Efficacy and Outcome Expectations in Person-Behavior-Outcome Dynamics*



Bandura highlights a distinction regarding expectations. As illustrated in [Figure 1](#), Bandura (1977) states that self-efficacy expectations mediate the relationship between the individual and behavior, while outcome expectations mediate the relationship between behavior and outcomes. Outcome expectation refers to an individual's estimation that a specific behavior will lead to a particular outcome. Self-efficacy expectation, on the other hand, reflects the individual's belief in their ability to perform the behaviors necessary to produce the desired outcome. Bandura (1997) further asserts that even if an individual believes that a sequence of actions will lead to specific outcomes, this belief will not influence their behavior if they doubt their ability to perform the required actions.

In his book introducing the concept of self-efficacy, Bandura (1997) also discusses the potential effects of the concept on teachers' instructional practices. In this regard, the researcher emphasizes that the evidence suggests teachers' beliefs about their instructional efficacy partially determine how they structure academic activities in their classrooms and shape their evaluations of students' intellectual abilities. For instance, teachers with low perceptions of instructional efficacy may believe that there is little they can do if students are poorly motivated and that their influence on students' intellectual development is severely constrained by adverse environmental factors. Consequently, Bandura highlights that teachers who strongly believe in their ability to foster learning are more likely to create positive learning experiences for their students. In contrast, those who doubt their instructional efficacy are likely to establish classroom environments that undermine students' confidence in their abilities and cognitive development.

All of this leads to the following conclusion: A teacher may believe that teaching the language of mathematics to students can yield positive outcomes for their learning experiences (outcome expectation). However, if the teacher harbors doubts about their ability to teach this skill effectively (self-efficacy expectation), they are unlikely to take action in this direction. Therefore, it is essential for teachers to possess a reasonable level of self-efficacy regarding their ability to impart this skill to their students.

2.3. Problem Statement

The main problem addressed in the study is to examine the psychometric properties of the scale developed to determine the self-efficacy of prospective mathematics teachers in teaching the language of mathematics in their future professional lives. Within this framework, the sub-problems to be addressed in the study are listed below:

- Is the category structure adopted in the scale appropriate?
- Are the person and item reliability and discrimination indices above the recommended thresholds?
- Are the fit statistics of the items within acceptable limits?
- Does the scale adequately target the sample?
- How are the items distributed?
- Are there any items that exhibit differential item functioning?
- Does the scale include a dimension other than the Rasch dimension?
- Does the scale meet the criterion of local independence?

3. Method

Investigations conducted to explore an existing situation and obtain detailed information about its characteristics are classified under the category of descriptive research (Çepni, 2007). The survey method, which is classified under this category, aims to collect data to gain information about a particular attribute present in individuals (Büyüköztürk *et al.*, 2019). In this study, conducted using the survey method, the psychometric properties of a scale developed to determine prospective teachers' self-efficacy in teaching the language of mathematics in their future professional lives were examined. Ethical approval for the study was obtained from the Scientific Research and Publication Ethics Committee of Trabzon University for Social and Human Sciences.

3.1. Participants

The participants of the study consisted of prospective teachers enrolled in the final semester of the elementary mathematics teacher education program at the universities listed in [Table 1](#). The selection of these universities was based on their geographical distribution across different regions and their heterogeneity in terms of admission scores. This approach aimed to enhance the diversity of the sample. It is considered that the various courses taken by students during their university education contribute to the development of the quality targeted by the scale. For this reason, it was deemed appropriate to select students in their final semester as the sample.

Seçer (2018) suggests that for a pilot study aimed at examining the compatibility of items with the scale, a group of 50 participants is sufficient if the number of items is fewer than 30. For the main application phase of the scale development process, Tabachnick and Fidell (2013) recommend a minimum of 300 participants, while Nunnally (1978) suggests a sample size ten times the number of items. The pilot study for the scale was conducted with 68 students, while the main application was completed with 457 students.

In the initial examination, it was identified that 33 students responded to the control item, which instructed them to select the "no comment" option, with answers other than "no comment." These students were excluded from the dataset. Additionally, five students who scored full marks on the scale were also removed from the dataset. The distribution of the remaining 419 students according to categorical variables is presented in [Table 1](#).

Table 1. *The Distribution of the Sample According to Categorical Variables*

	University														Gender	
	Trabzon	Bayburt	İstanbul Medipol	Yozgat Bozok	Recep Tayyip Erdoğan	Sakarya	Sivas Cumhuriyet	Dokuz Eylül	Düzce	Manisa Celal Bayar	Necmettin Erbakan	Giresun	Erzincan Binali Yıldırım	Kilis 7 Aralık	Female	Male
n	62	28	16	21	35	25	24	49	38	10	29	29	17	36	330	89

3.2. The Development Process of the Scale Form

In the study, a literature review was first conducted to establish a foundation for defining the language of mathematics and identifying its components. Based on the findings obtained from this review, an item pool was created. Considering studies in the literature reporting that negatively worded items do not function as intended (Boone *et al.*, 2014; Carlson *et al.*, 2011; Suárez-Alvarez *et al.*, 2018; Zeng *et al.*, 2020), no negatively worded items were included in the item pool.

After the item pool was created, a group discussion was initially conducted with 9 prospective teachers regarding the items. In this discussion, one of the researchers read each item individually with the prospective teachers and asked them to express their opinions on the clarity and meaning of each item. Based on the feedback received, changes were made to some items. Following this process, to enhance the validity of the scale from various perspectives, the scale items were shared online with 18 faculty members (6 Professors, 11 Associate Professors, 1 Assistant Professor) working in the field of mathematics education at 13 different universities, and they were asked to evaluate the items. As a result of the improvements made based on the feedback, the draft scale was prepared for the pilot application. After the pilot application was conducted, the finalized scale, consisting of 20 items (See [Appendix A](#)), was made ready for the main application.

3.3. The Administration Process of the Scale

The scale was administered to participants online via the Google Forms platform. Upon clicking the link provided to access the scale, participants were presented with an invitation letter and an option to confirm their voluntary participation in the study. Those who chose to participate voluntarily were asked to indicate their gender and the university they were attending. Following this, an explanatory text regarding the scale was provided to the participants.

Figure 2. *Explanations Provided Alongside the Scale Items*

Mathematical Notation: Forms that represent mathematical concepts. For example: $\int f(x)dx$, $\frac{a}{b}$, m_d , etc.

Mathematical Symbol: Forms used to represent mathematical relationships or operations. For example: $\leq$, $\neq$, $\div$, etc.

a)

Mathematical Term: Words specific to mathematics. For example: denominator, multiplicative inverse, angle bisector, chord, remainder, quotient, tenths place, etc.

b)

The scale items included five categories from which participants could select only one: very low proficiency, low proficiency, moderate proficiency, proficient, and very proficient. Items containing mathematical notation/symbol or mathematical term expressions were presented along with the explanations shown in [Figures 2a](#) and [2b](#), respectively.

To determine whether participants were responding to the scale items randomly, a control item was placed as the tenth item. In this control item, participants were instructed to select only the "no comment" option provided. Finally, when scoring participants' responses, the framework adopted was that higher scores indicated higher self-efficacy.

3.4. Data Analysis

3.4.1. Rating Scale Model

Rating scales are widely used for measuring psychological constructs. However, directly subjecting raw scores obtained from these scales to statistical analyses to answer research questions brings several limitations (Boone *et al.*, 2014). One of these limitations is that the data obtained may not meet the normality assumption, which is a prerequisite for parametric tests. In such a scenario, interpretations based on the means of the scores can be misleading (Neumann *et al.*, 2011).

Additionally, in such scales, participants are often presented with more than two categories to choose from, and responses to these categories are scored using consecutive integers. This practice assumes that consecutive categories represent equal differences in the amount of the latent variable being measured. Furthermore, if the same category is selected for two different items, it is assumed that both contribute equally to the total score, as the raw scores from both items are treated as having the same impact on the amount of the latent variable. These assumptions are based on the premise that the data obtained from the scale are interval-level data.

Researchers adopting the RMT argue that this common practice is flawed, suggesting that both the categories and the items may contribute differently to the amount of the latent variable (Boone *et al.*, 2014). In other words, they propose that the difference between two individuals scoring 20 and 30 on a scale may not be equivalent to the difference between two individuals scoring 30 and 40. Therefore, they advocate treating raw scores obtained from rating scales as ordinal data rather than interval data (Bond & Fox, 2007).

Rasch models provide significant advantages in the process of scale development and the quantification of responses to scales (Bond & Fox, 2007). One of these advantages is the transformation of raw scores, treated as ordinal data, into interval-level data (Boone *et al.*, 2014). Additionally, Smith *et al.* (2002) outline several other advantages of RMT compared to Classical Test Theory (CTT), including offering solutions to the problem of data loss, providing validity and reliability estimates for both item and person measures, focusing on individuals and items rather than group statistics, and positioning the measures of persons and items on an equal-interval scale in the same unit of measurement. The fact that the measures of items and persons are expressed in the same unit, referred to as logits, enables comparisons both within and between items and persons. This, in turn, provides educators with a richer set of information, allowing for more detailed interpretations based on the results of their measurements.

Within RMT, the Rating Scale Model[†] (RSM) is utilized for the analysis of rating scales. The functioning of the RSM is based on the mathematical equation shown below. In this equation, the variable F_j represents the threshold values at which the probabilities of selecting two

[†] An explanation of the model can be found in the research report by Çekmez *et al.* (2023)

consecutive categories are equal. As implied by its definition, the number of threshold values is one less than the number of categories. For instance, in the context of this study, F_4 represents the threshold value at which the probabilities of selecting the categories "very proficient" and "proficient" are equal. Among the other variables, B_n represents the measure of the n -th person, D_i represents the measure of the i -th item, and P_{nij} and $P_{ni(j-1)}$ represent the probabilities of selecting consecutive categories.

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

3.4.2. Functionality of the Category Structure

The validity of measurements obtained through rating scale-based assessments primarily depends on whether the category structure adopted in the scale functions as intended by the researchers. This is because the estimates derived from the measurement are based on the quantification of responses given to these categories. It is essential to consider the proper functioning of the category structure in a designed scale not as an assumption but as a condition that must be tested (Boone *et al.*, 2014). The rationale behind this approach is that respondents' reactions to items may differ from the expectations of the test designers (Andrich, 1996).

Participants' responses to the items were scored using integers ranging from 1 to 5, increasing from the "very low proficiency" category to the "very proficient" category. As a natural consequence, the average measures of participants endorsing any given category in an item should be greater than the average measures of participants endorsing a lower-scoring category. Additionally, to test the functionality of the categories, Linacre (2002) proposed the following six criteria: (i) category mean measures should increase monotonically, (ii) all categories should be endorsed at least 10 times, (iii) outfit statistics should not exceed 2, (iv) category threshold measures should increase monotonically, (v) the difference between successive thresholds should be greater than 1.4 logits but less than 5 logits, and (vi) each category probability curve should peak. These criteria were taken into account when evaluating the functionality of the categories.

3.4.3. Distribution of Items and the Targeting Adequacy of the Scale

One of the factors influencing the construct validity of a scale is its ability to include items that correspond to different levels of the attribute it aims to measure (Boone *et al.*, 2014). If the items are confined to a narrow range of measurements, they may fail to adequately capture the low, medium, and high levels of the attribute being assessed. Within the framework of Rasch analysis, one of the tools used to evaluate the hierarchy of scale items and their ability to address different levels of the measured attribute is the person-item map (also referred to as the Wright map). The person-item map positions individuals and items on an equal-interval scale in terms of a measurement unit called logits. Another function of this map is to provide an opportunity to assess the targeting adequacy of the scale for the sample to which it is applied (Boone *et al.*, 2014).

The targeting efficacy of a scale for the sample to which it is applied is another factor influencing its validity. A well-targeted scale means that the items are neither too difficult nor too easy for the individuals in the sample (Boone *et al.*, 2014). One indicator of good targeting is the alignment between the regions where individuals are concentrated and the regions where items are concentrated on the person-item map (Bond & Fox, 2007). In such cases, the error in predictions for individuals will decrease, and it will be possible to distinguish individuals from one another with greater precision. A difference in the absolute value between the mean of the persons and the mean of the items smaller than 0.5 indicates

slight misalignment, while a difference greater than 1 indicates significant misalignment (Duncan *et al.*, 2003). These considerations were taken into account when evaluating the appropriateness of the distribution of the scale items and its targeting adequacy.

3.4.4. Item Fit Statistics

In the context of RMT, fit refers to the extent to which the data obtained from a measurement process align with the expectations of the model (Bond *et al.*, 2021). Items that exhibit a high degree of misfit with the model's expectations do not contribute to the measurement of the intended construct (Oon & Subramaniam, 2011).

The reporting of fit statistics is typically conducted using mean-square residuals. The first of these statistics, infit, is sensitive to unexpected responses to items that are close to an individual's ability level. The second statistic, outfit, is sensitive to unexpected responses to items that are far from an individual's ability level (Bond & Fox, 2007). A Rasch analysis provides fit indices for both individuals and items. However, since the invariance of item measures is a critical requirement in a scale, item fit is prioritized over person fit (Wright & Linacre, 1994). Wright and Linacre (1994) recommend that fit indices for rating scales should fall within the range of 0.6 to 1.4.

Another indicator used to determine whether the items in a scale contribute to the measurement of the intended construct is the point-measure correlation (Boone & Staver, 2020). A correlation value close to zero or negative indicates that the item moves in opposition to the overall scale (Bond *et al.*, 2021). The proposed lower threshold for point-measure correlations is 0.3 (Boone, 2020). These considerations were taken into account when evaluating whether the items in the scale serve its intended purpose.

3.4.5. Reliability and Discrimination

In the context of RMT, reliability coefficients are interpreted similarly to those in CTT (Linacre, 2022). For this reason, the closer the coefficients are to 1, the higher the internal consistency of the measurement process is considered to be. Unlike CTT, however, RMT provides reliability coefficients for both persons and items. The reliability coefficient for items indicates the potential of the scale items to define a hierarchy based on the measured construct (Linacre, 2022). A high person reliability index, on the other hand, suggests that the scale allows for both high and low scores and that the calculated person measures are consistent (Bond *et al.*, 2021).

Another concept directly related to reliability in Rasch analysis is the separation index (G). This index is calculated by dividing the standard deviation of the true scores by the standard deviation of the errors (Linacre, 2022). Thus, it reflects the distribution of the sample based on measurement precision (Wright & Stone, 2004). The separation index calculated for persons provides information about the extent to which individuals can be classified (Bond *et al.*, 2021). A person separation index greater than 3 is considered excellent, between 2 and 3 is good, and a value of 1.5 is the threshold for acceptability (Duncan *et al.*, 2003). For the item separation index, a minimum value of 1.5 is required for group-level analysis, while a value of at least 2.5 is necessary for individual-level analysis (Tennant & Conaghan, 2007). Based on the results of a measurement, the number of distinct strata into which the sample can be divided is calculated using the person separation index with the following formula (Linacre, 2022).

$$\text{Strata} = \frac{4G + 1}{3}$$

These considerations were taken into account when evaluating the internal consistency demonstrated by the scale items and their ability to distinguish between individuals.

3.4.6. Unidimensionality and Local Independence

One of the fundamental assumptions of RMT is the principle of unidimensionality, which states that a measurement tool should serve to measure only a single latent variable (Linacre, 2022). To test whether a measurement tool meets this principle, Principal Component Analysis of Residuals (PCAR) is used (Boone & Staver, 2020). If the eigenvalue of the first principal component calculated from the correlation matrix of the residuals is less than 2, it is concluded that the variance not explained by the latent variable is random noise (Boone & Staver, 2020). Additionally, if the variance explained by the latent variable accounts for more than half of the total variance, this is considered another indicator of unidimensionality.

The second fundamental assumption of RMT is the principle of local independence (Bond *et al.*, 2021). This principle asserts that responses to one item should not depend on responses to another item (Fan & Bond, 2019). If items are independent, there should be no high correlations among the residuals of the items (Fan & Bond, 2019). To suspect dependency between two items, Linacre (2022) suggests that the correlation coefficient between their residuals should exceed .7. Based on these considerations, a PCAR analysis was conducted, and the correlation coefficients among the residuals were calculated to determine whether the scale violated these two assumptions.

3.4.7. Differential Item Functioning

Whether the items in a scale function similarly across different subgroups of the sample is tested through differential item functioning (DIF) analysis (Linacre, 2022). The invariance of scale items is a critical factor influencing the construct validity of the scale. This principle of invariance implies that the ranking of items in terms of their endorsability should not differ across subgroups of the sample (Boone *et al.*, 2014).

To test the invariance of the scale items, the two largest subgroups within the sample, based on the university variable, were identified, and DIF analysis was conducted using the Winsteps software. A scale item is considered to exhibit bias between the two subgroups if the absolute difference between the item measures in the two groups exceeds .64 and the probability of this difference occurring is less than .05 (Boone *et al.*, 2014).

4. Findings

In this section, the findings obtained from the analysis of the data collected during the main application are presented under headings aligned with the research questions.

4.1. Findings Regarding the Functionality of the Category Structure

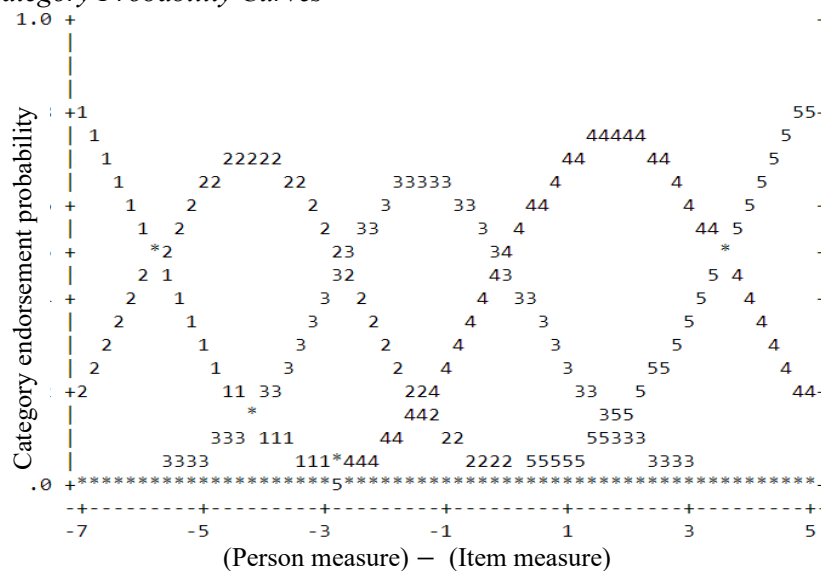
In the examination of the mean measures of individuals endorsing different categories, no findings contrary to expectations were observed. However, a misfit was identified for the 1-point and 2-point categories in items coded M4 and M7. In both items, only one individual endorsed the 1-point category, while 40 and 17 individuals, respectively, endorsed the 2-point category. Therefore, the observed misfit for both items arises from the comparison of the measure of one individual with the measures of 40 and 17 individuals, respectively.

The findings related to the first five criteria proposed by Linacre (2002), as outlined in the data analysis section, are presented in [Table 2](#). The results indicate that all five criteria were met.

Table 2. Findings Related to the First Five Criteria

Category (Point)	Frequency (%)	Mean measure	Outfit (MnSq)	Category threshold
Very low proficiency (1)	80 (1)	−4.04	1.51	—
Low proficiency (2)	671 (8)	−2.67	1.09	−5.75
Moderate proficiency (3)	2394 (29)	−0.38	0.91	−2.68
Proficient (4)	4184 (49)	1.4	0.95	−0.03
Very proficient (5)	1051 (13)	2.91	1.04	3.54

To evaluate the final criterion, category probability curves were generated, as shown in Figure 3. In the coordinate system presented in Figure 3, the vertical axis represents the probability value, while the horizontal axis represents the difference between the person measure and the item measure. This coordinate system illustrates the probability of a participant endorsing a specific category for any given item based on the difference between the participant's measure and the item's measure. As observed, each category has a range where its selection probability is the highest. In Linacre's terms, all category curves have formed peaks.

Figure 3. Category Probability Curves

4.2. Findings Regarding the Distribution of Items and the Targeting Adequacy

The person-item map resulting from the measurement is presented in Figure 4. The vertical line in the center of the figure represents the equal-interval scale on which the measures of persons and items are displayed. On the left side of the figure, the distribution of the measures of individuals is shown. The “#” symbol represents 4 individuals, while the “.” symbol represents 1, 2, or 3 individuals. The letters M, S, and T on both the left and right sides of the figure indicate the mean, 1 standard deviation, and 2 standard deviations, respectively.

In the person section, moving upward indicates an increase in participants' self-efficacy measures. In other words, as one moves upward, the raw scores obtained increase. On the right side, moving upward indicates that the items become more difficult to endorse. In other words, the total raw score obtained for the item decreases as one moves upward. Considering the person-item map generated in the study, it is observed that the persons and items are positioned in relation to each other. In a Rasch analysis, the mean of the item measures is calibrated to a default value of 0 logits. In this study, the mean of the person measures was found to be 0.7 logits.

Figure 4. *Person-Item Map*

4.3. Findings Related to Item Fit Statistics

The fit statistics and point-measure correlation values are presented in Table 3. As seen in Table 3, the outfit statistics of the items range between 0.61 and 1.3, the infit statistics range between 0.63 and 1.3, and the point-measure correlations vary between .48 and .76.

Table 3. *Measures, Fit Statistics, and Point-Measure Correlation Values of the Items*

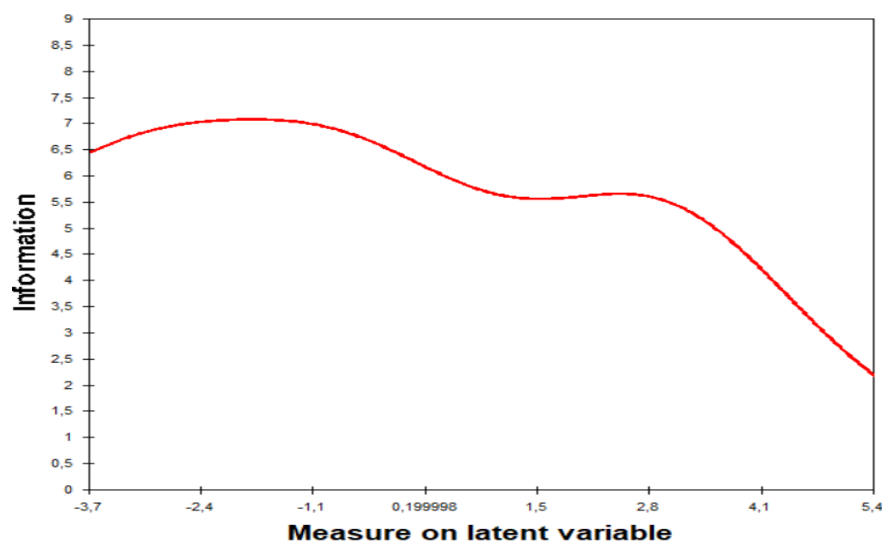
Item	Item measure (SE)	Outfit (MnSq)	Infit (MnSq)	Pmc	Item	Item measure (SE)	Outfit (MnSq)	Infit (MnSq)	Pmc
M1	3.83 (0.08)	0.61	0.63	.76	M11	-0.50 (0.09)	1.03	1.05	.65
M2	3.32 (0.08)	0.64	0.67	.73	M12	-0.64 (0.09)	1.05	1.06	.63
M3	1.22 (0.08)	0.82	0.82	.74	M13	-0.72 (0.09)	1.23	1.26	.63
M4	0.72 (0.09)	1.26	1.23	.60	M14	-0.76 (0.09)	0.86	0.88	.66
M5	0.28 (0.09)	1.06	1.06	.60	M15	-0.82 (0.09)	1.23	1.25	.61
M6	0.07 (0.09)	1.29	1.30	.62	M16	-0.83 (0.09)	1.17	1.20	.54
M7	-0.10 (0.09)	0.77	0.79	.70	M17	-0.91 (0.09)	0.93	0.96	.62
M8	-0.14 (0.09)	0.75	0.76	.70	M18	-0.73 (0.09)	0.93	0.85	.48
M9	-0.36 (0.09)	0.91	0.92	.67	M19	-0.99 (0.09)	1.30	1.19	.49
M10	-0.45 (0.09)	1.08	1.11	.61	M20	-1.49 (0.09)	1.13	1.11	.61

Pmc: Point-measure correlation; SE: Standart error;

4.4. Findings Related to Reliability and Discrimination

The reliability coefficient (Cronbach's alpha) of the raw scores was calculated as .92, the reliability coefficient of the persons' logit measures as .91, and the reliability coefficient of the items' logit measures as .99. These coefficients exceed the threshold of .7 recommended in the literature (Boone *et al.*, 2014). In the current measurement, the person logit measures range from -3.68 to 5.38, with six individuals exhibiting measures greater than 4 logits. The graph of the test information function was generated for this continuum to determine the precision with which the test measures individuals at varying ability levels. According to the relationship between the standard error of the measurement and the value of the information function, a standard error of 0.5 is equivalent to an information value of 4 (Boone & Staver, 2020). As illustrated in Figure 5, the test measured almost all individuals in the sample with a standard error of less than 0.5.

Figure 5. Test Information Function



Additionally, the item discrimination index was calculated as 14.34, and the person discrimination index as 3.19. The strata value calculated based on the person discrimination index was approximately 4.

4.5. Findings Regarding Unidimensionality and Local Independence

As a result of the PCAR analysis conducted in this study, the variance explained by the latent variable accounted for approximately 57.1% of the total variance, while the eigenvalue of the first principal component calculated from the residual correlation matrix was found to be 1.91. On the other hand, for the assumption of local independence, the correlation coefficients among item residuals were calculated and found to range between $-.25$ and $.41$.

4.6. Findings from the Differential Item Functioning Analysis

The two largest subgroups within the sample, in terms of the university variable, are Trabzon University and 9 Eylül University. To determine whether the items exhibit bias, the hypothesis "Item measures are equal across the two groups" was tested for each item, considering these two universities. The findings obtained from the DIF analysis are presented in Table 4. As seen in Table 4, the absolute differences between the measures of all items across the two groups are less than 0.64, and the p -values obtained for all items are greater than the significance threshold (.05).

Table 4. *DIF Analysis Findings for Trabzon and 9 Eylül Universities Subgroups*

Item	Trabzon		9 Eylül		Difference	<i>t</i>	<i>p</i>
	DIF measure	SE	DIF measure	SE			
M1	3.73	0.20	3.69	0.22	0.04	0.14	.89
M2	3.47	0.19	3.18	0.21	0.29	0.99	.33
M3	1.31	0.19	1.22	0.22	0.09	0.31	.76
M4	0.77	0.20	1.19	0.20	0.42	−1.42	.16
M5	0.22	0.20	0.28	0.23	0.06	−0.21	.83
M6	0.18	0.20	−0.12	0.23	0.30	0.97	.33
M7	0.22	0.20	0.09	0.23	0.13	0.42	.68
M8	−0.41	0.22	0.09	0.23	0.50	−1.61	.11
M9	−0.03	0.20	−0.56	0.24	0.53	1.7	.09
M10	−0.62	0.21	−0.28	0.23	0.34	−1.08	.28
M11	−0.75	0.21	−0.28	0.33	0.47	−1.49	.14
M12	−0.71	0.21	−0.96	0.24	0.25	0.79	.43
M13	−1.02	0.21	−0.78	0.24	0.32	−0.72	.47
M14	−0.79	0.21	−0.67	0.21	0.12	−0.39	.70
M15	−1.02	0.20	−0.90	0.24	0.12	−0.36	.72
M16	−0.53	0.21	−0.67	0.24	0.14	0.44	.66
M17	−0.84	0.21	−1.19	0.24	0.35	1.10	.27
M18	−0.73	0.21	−0.90	0.24	0.17	0.53	.60
M19	−0.99	0.21	−1.13	0.24	0.14	0.45	.66
M20	−1.52	0.22	−1.37	0.25	0.15	−0.44	.66

5. Discussion, Conclusion, and Suggestions

One of the aspects examined in terms of validity in this study was the functionality of the adopted category structure in responding to the scale items. The findings revealed that the proposed criteria regarding the frequency of categories, mean measures, fit indices, category threshold values, and probability curves of the categories were met. Consequently, since the adopted category structure demonstrated good functionality in terms of measurement, it is recommended that the scale be used with its current category structure in future studies.

Another issue addressed in the study was the examination of the measurement in terms of reliability and discrimination. The reliability coefficient calculated based on the raw scores of individuals indicates that the measurement has a high level of internal consistency. The reliability coefficients calculated based on the logit measures of both persons and items were found to be higher than the thresholds suggested in the literature. This suggests that if the scale were applied to a different sample in the future, the item measures would likely be close to the measures obtained in this study (Bond *et al.*, 2021). Furthermore, the item discrimination value was calculated to be above the threshold recommended in the literature. Accordingly, the strata value for the items was also found to be quite high. These values indicate that the scale items, as a whole, can measure a wide range of the self-efficacy variable related to teaching the language of mathematics. Additionally, the strata value for persons, calculated based on the person discrimination index, was approximately 4. This value suggests that the scale can classify persons into four levels in terms of their possession of the relevant attribute.

When the distribution of the scale items on the person-item map is considered, it is observed that the items are distributed across a wide range of the measured variable. However, there is a significant gap between the second most difficult item to endorse and the third item on the

scale. This indicates that the scale may lack sensitivity in distinguishing individuals whose measures fall within this range. Therefore, enriching the scale with an item or items that fall within this range should be addressed in future research.

In the measurement conducted, the mean of the individuals' measures was calculated to be 0.7 logits higher than the mean of the items' measures, indicating a moderate level of mismatch in terms of targeting the sample. Thus, the overall endorsement of the scale items is slightly easier for the sample. Implementing the aforementioned enrichment would also positively contribute to the scale's targeting adequacy.

In terms of fit, both fit statistics for the items fall within the recommended range. Therefore, the responses to the items did not deviate significantly from the model's expectations. This indicates that the items collectively serve the measurement of the targeted attribute (Oon & Subramaniam, 2011). Additionally, the point-measure correlations calculated for the items consist of high positive values. These values suggest that each item on the scale behaves in the same direction and is thus related to the intended attribute to be measured.

For the two fundamental assumptions of RMT, unidimensionality and local independence, PCAR analysis was conducted, and the residual correlation matrix was examined. The results of the PCAR analysis showed that the proportion of variance explained by Rasch measures in the total variance and the eigenvalue of the first principal component calculated from the residual correlation matrix indicate that the scale does not contain another dimension apart from the Rasch dimension. Therefore, the developed scale serves the measurement of a single latent variable. Additionally, since the correlation coefficient among the residuals is smaller than the recommended upper threshold, the scale is considered not to violate the local independence criterion.

To test whether the items exhibit bias, DIF analysis was conducted by considering the two largest groups in the sample in terms of the university variable. The differences between the item measures calculated for both groups were found to be below the threshold recommended in the literature. This indicates that the scale items behave similarly across different samples, meaning that their positions on the person-item map do not show significant differences.

Considering the findings obtained from the study and the criteria recommended in the literature, it is concluded that the scale developed in this study performs a valid and reliable measurement at an acceptable level. Consequently, the scale can be used in studies aiming to determine the self-efficacy of primary mathematics teacher candidates regarding teaching the language of mathematics.

Ethics Committee Approval

This study was carried out with the approval of the Scientific Research and Publication Ethics Committee of Trabzon University for Social and Human Sciences, granted by the decision dated 24.04.2024 and numbered E-81614018-050.04-2400019593.

Conflict of Interest

The authors declare that they have no conflict of interest.

Author Contribution

Erdem Çekmez: Literature review, data collection, development of the conceptual framework, writing of the research report, data analysis, and interpretation. **Mustafa Güler:** Literature review, data collection, development of the conceptual framework, writing of the research report. **Taha Ertuğrul Kuzu:** Literature review, development of the conceptual framework.

Orcid

Erdem Çekmez  <https://orcid.org/0000-0001-8684-2820>

Mustafa Güler  <https://orcid.org/0000-0002-4082-7585>

Taha Ertuğrul Kuzu  <https://orcid.org/0000-0001-7422-0231>

REFERENCES

- Açıl, E., & Zeybek, Z. (2017). Students' ability to use and understand mathematical language with their teacher's ability to notice the ways of how students use mathematical language? *Pamukkale University Journal of Education*, 42, 87-107.
- Adler, J. (1999). The dilemma of transparency: Seeing and seeing through talk in the mathematics classroom. *Journal for Research in Mathematics Education*, 30(1), 47-64. <https://doi.org/10.2307/749629>
- Akcan, F. (2023). *The effect of the online supportive learning tool developed for digital security proficiency in social studies course on students' self-efficacy perceptions* (Publication No: 776088) [Doctoral dissertation, Gazi University]. Council of Higher Education Thesis Center.
- 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>
- Araujo, Z. d., Roberts, S. A., Willey, C., & Zahner, W. (2018). English learners in K–12 mathematics education: A review of the literature. *Review of Educational Research*, 88(6), 879-919. <https://doi.org/10.3102/0034654318798093>
- Baber, R. L. (2011). *The language of mathematics: Utilizing math in practice*. John Wiley & Sons.
- Bali-Çalikoğlu, G. (2002). Language in mathematics teaching scale. *Hacettepe University Journal of Education*, 23, 57-61.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191-215. <https://doi.org/10.1037/0033-295X.84.2.191>
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman and Company.
- Barton, B. (2008). *The language of mathematics: Telling mathematical tales*. Springer. <https://doi.org/10.1007/978-0-387-72859-9>
- Barwell, R. (2016). Formal and informal mathematical discourses: Bakhtin and Vygotsky, dialogue and dialectic. *Educational Studies in Mathematics*, 92(3), 331-345. <https://doi.org/10.1007/s10649-015-9641-z>
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Bond, T. G., Yan, Z., & Heene, M. (2021). *Applying the Rasch model: Fundamental measurement in the human sciences* (4th ed.). Routledge
- Boone, W. J. (2020). Rasch basics for the novice. In M. S. Khine (Ed.), *Rasch measurement: Applications in quantitative educational research* (pp. 9-30). Springer. <https://doi.org/10.1007/978-981-15-1800-3>
- Boone, W. J., & Staver, J. R. (2020). *Advances in Rasch analysis in the human sciences*. 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*. Springer. https://doi.org/10.1007/978-94-007-6857-4_4
- Burns, M. (2006). Marilyn Burns on the language of math. *Instructor*, 115(7), 41-43.

- Büyüköztürk, Ş., Kılıç-Çakmak, E., Akgün, Ö., Karadeniz, Ş., & Demirel, F. (2019). *Bilimsel araştırma yöntemleri [Scientific research methods]*. 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. *Psychological Assessment*, 23(2), 582-562. <https://doi.org/10.1037/a0022484>
- Çekmez, E., Tuti, G., & Terzi, S. Y. (2023). Investigation of the psychometric properties of a mathematics anxiety test for 3rd and 4th grade primary school students. *Anadolu University Journal of Education Faculty*, 7(4), 819-841. <https://doi.org/10.34056/aujef.1254390>
- Çepni, S. (2007). *Araştırma ve proje çalışmalarına giriş [Introduction to research and project studies]*. Celepler Matbaacılık.
- 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)
- Durkin, K., & Shire, B. (1991). Lexical ambiguity in mathematical contexts. In K. Durkin & B. Shire (Eds.), *Language in mathematical education* (pp. 71-84). Open University Press.
- Erath, K., Prediger, S., Quasthoff, U. M., & Heller, V. (2018). Discourse competence as important part of academic language proficiency in mathematics classrooms: The case of explaining to learn and learning to explain. *Educational Studies in Mathematics*, 99(2), 161-179. <https://doi.org/10.1007/s10649-018-9830-7>
- 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: Fundamental techniques* (pp.83-102). Routledge. <https://doi.org/10.4324/9781315187815>
- Gökmen, A.D.Y., & Ulusoy, M. (2022). The role of vocabulary knowledge and comprehension skills on retelling skills: Different path analysis tests. *Pamukkale University Journal of Education*, 56, 372-400. <https://doi.org/10.9779/pauefd.1026827>.
- Gürefe, N. (2018). Mathematical language skills of mathematics prospective teachers. *Universal Journal of Educational Research*, 6(4), 661-671. <https://doi.org/10.13189/ujer.2018.060410>
- Gray, V. (2004). *The language of mathematics: A functional definition and development of an instrument to measure teacher perceived self-efficacy* [Doctoral dissertation, Oregon State University]. <http://hdl.handle.net/1957/31694>
- Halliday, M. A. K. (1978). *Language as social semiotic*. Edward Arnold.
- Kabael, T., & Yayan, B. (2017). Effect of self-evaluation on pre-service mathematics teachers' self-efficacy in language of mathematics. *Anadolu Journal of Educational Sciences International*, 7(1), 1-34. <https://doi.org/10.18039/ajesi.292575>
- Kuzu, T. (2023). Multilingual meaning making – An explorative study of German-Turkish learners' translanguaging processes regarding the part-whole-concept. *Journal für Mathematik-Didaktik*, 44(4), 325-353. <https://doi.org/10.1007/s13138-023-00219-z>
- Kuzu, T., & Ratzke, C. (2024). Explanatory videos in the context of learning processes: An interdisciplinary interpretative interaction analysis of production and reflection processes. *Journal of Pedagogical Research*, 8(2), 227-259. <https://doi.org/10.33902/JPR.202426555>
- Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3(1), 85-106.
- Linacre, J. M. (2022). *A user guide to Winsteps Ministep Rasch model computer programs*. <https://www.winsteps.com/a/Winsteps-Manual.pdf>

- Millî Eğitim Bakanlığı [Ministry of National Education]. (2018a). *Matematik dersi öğretim programı (İlkokul ve ortaokul 1,2,3,4,5,6,7, ve 8. sınıflar)* [Mathematics curriculum (Primary and secondary school grades 1, 2, 3, 4, 5, 6, 7, and 8)]. Millî Eğitim Bakanlığı.
- Millî Eğitim Bakanlığı [Ministry of National Education]. (2018b). *Ortaöğretim matematik dersi (9, 10, 11 ve 12. sınıflar) öğretim programı* [Secondary education mathematics curriculum (Grades 9, 10, 11, and 12)]. Millî Eğitim Bakanlığı.
- Millî Eğitim Bakanlığı [Ministry of National Education]. (2024a). *Ortaöğretim matematik dersi öğretim programı (Hazırlık, 9, 10, 11 ve 12. sınıflar)* [Secondary Education Mathematics Curriculum (Preparatory, Grades 9, 10, 11, and 12)]. Millî Eğitim Bakanlığı.
- Millî Eğitim Bakanlığı [Ministry of National Education]. (2024b). *Ortaokul matematik dersi öğretim programı (5, 6, 7 ve 8. sınıflar)* [Middle school mathematics curriculum (Grades 5, 6, 7, and 8)]. Millî Eğitim Bakanlığı.
- Moleko, M. M. (2021). Teachers' perspectives on addressing linguistic factors affecting visualization of mathematics word problems. *Eurasia Journal of Mathematics, Science and Technology Education*, 17(11), em2029. <https://doi.org/10.29333/ejmste/11248>
- Moore, R. C. (1994). Making the transition to formal proof. *Educational Studies in Mathematics*, 27(3), 249-266. <https://doi.org/10.1007/bf01273731>
- Morgan, C. (2020). Mathematical language. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 540-543). Springer. <https://doi.org/10.1007/978-3-030-15789-0>
- Moschkovich, J. (2015). Academic literacy in mathematics for English learners. *The Journal of Mathematical Behavior*, 40, 43-62. <https://doi.org/10.1016/j.jmathb.2015.01.005>
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. NCTM.
- 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. <https://doi.org/10.1080/09500693.2010.511297>
- Nunnally, J. C. (1978). *Psychometric theory*. (2nd ed.). McGraw-Hill.
- 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 (Ed.), *Applications of Rasch measurement in learning environments research* (pp. 119-139). Sense Pub.
- Özgen, K., & Bindak, R. (2008). The development of self-efficacy scale for mathematics literacy. *Kastamonu Eğitim Dergisi*, 16(2), 517-528.
- Peng, P., Lin, X., Ünal, Z., Lee, K., Namkung, J., Chow, J., ... & Sales, A. (2020). Examining the mutual relations between language and mathematics: A meta-analysis. *Psychological Bulletin*, 146(7), 595-634. <https://doi.org/10.1037/bul0000231>
- Pimm, D. (1987). *Speaking Mathematically: Communication in mathematics classrooms*. Routledge.
- Prediger, S. (2019). Investigating and promoting teachers' expertise for language-responsive mathematics teaching. *Mathematics Education Research Journal*, 31, 367-392. <https://doi.org/10.1007/s13394-019-00258-1>
- Prediger, S., Kuzu, T., Schüler-Meyer, A., & Wagner, J. (2019). One mind, two languages – separate conceptualizations? A case study of students' bilingual modes for dealing with language-related conceptualizations of fractions. *Research in Mathematics Education*, 21(2), 188-207. <https://doi.org/10.1080/14794802.2019.1602561>
- Riccomini, P. J., Smith, G. W., Hughes, E. M., & Fries, K. M. (2015). The language of mathematics: The importance of teaching and learning mathematical vocabulary. *Reading & Writing Quarterly*, 31(3), 235-252. <https://doi.org/10.1080/10573569.2015.1030995>

- Schleppegrell, M. J. (2007). The linguistic challenges of mathematics teaching and learning: A research review. *Reading & Writing Quarterly*, 23(2), 139-159. <https://doi.org/10.1080/10573560601158461>
- Seçer, İ. (2018). *Psikolojik test geliştirme ve uyarlama süreci: SPSS ve LISREL uygulamaları (2. Baskı) [The process of developing and adapting psychological tests: SPSS and LISREL applications (2nd Edition)]*. Anı Yayıncılık.
- Smith Jr, E. V., Conrad, K. M., Chang, K., & 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., & Muñiz, J. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema*, 30(2), 149-158. <https://doi.org/10.7334/psicothema2018.33>
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using Multivariate Statistics* (6th ed.). Pearson.
- Tennant, A., & Conaghan, P. (2007). The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Care and Research*, 5(8), 1358–1362. <https://doi.org/10.1002/art.23108>
- Uptegrove, E. B. (2015). Shared communication in building mathematical ideas: A longitudinal study. *Journal of Mathematical Behavior*, 40, 106–130. <https://doi.org/10.1016/j.jmathb.2015.02.001>
- Wilkinson, L. C. (2018). Teaching the language of mathematics: What the research tells us teachers need to know and do? *The Journal of Mathematical Behavior*, 51, 167-174. <https://doi.org/10.1016/j.jmathb.2018.05.001>
- Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8(3), 370.
- Wright, B.D., & Stone, M.H. (2004). *Making measures*. Phaneron Press
- 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>

Appendix A. Scale Items

M5. I can provide explanations that help students distinguish between the meanings of words used both in everyday language and in mathematical language (e.g., similar, rational).

M18. I can identify whether students use mathematical terms appropriately.

M14. I can serve as a model for my students in writing mathematical expressions in accordance with the rules.

M16. I can ensure that the mathematical expressions my students write in their notebooks comply with the rules.

M20. I can help my students understand the purpose of using letters or symbols when forming equations.

M9. I can help students develop the habit of using mathematical terms appropriately.

M11. I can improve students' ability to verbally express texts written in mathematical language.

M15. I can help my students recognize the importance of graphs in mathematics.

M4. I can encourage my students to be willing to discuss their mathematical ideas with their classmates.

M8. I can help my students understand the relationship between different representations of mathematical concepts.

M10. I can organize the classroom environment to encourage students to share their mathematical ideas verbally.

M17. I can identify the mistakes students make when using mathematical language.

M7. I can improve students' ability to express mathematical knowledge in written form.

M13. I can help students develop graph-reading skills.

M6. I can choose words appropriate to students' age levels when explaining mathematical knowledge.

M3. I can design activities that improve students' ability to express relationships given in verbal problems mathematically.

M2. I know where to look for activities designed to improve students' mathematical language skills if I want to find them.

M19. I can serve as a model for my students in drawing geometric shapes by considering the measurements of the elements that make up the shape (e.g., if an angle measures more than 90 degrees, drawing it to appear as an obtuse angle).

M12. I pay attention to ensuring that my students correctly read mathematical symbols and notations.

M1. I am familiar with types of activities designed to assess students' mathematical language proficiency.