

FARKLI KURAMLAR KAPSAMINDA HESAPLANAN MADDE AYIRT EDİCİLİK PARAMETRELERİNİN KARŞILAŞTIRILMASI

THE COMPARISON OF ITEM DISCRIMANTION PARAMETERS ESTIMATED FROM DIFFERENT TEST THEORIES

Ezgi MOR DİRLİK¹

ÖZ: Bu çalışmanın temel amacı Klasik Test Kuramı(KTK) ve Parametrik Olmayan Madde Tepki Kuramına(POMTK) dayalı olarak kestirilen madde ayırt edicilik parametrelerinin karşılaştırılması ve benzerliklerinin incelenmesidir. Bu amaç doğrultusunda nicel araştırma yaklaşımı içinde yer alan korelasyonel türde bir araştırma yöntemi tercih edilmiştir. İlkokul ve ortaokullarda görev yapan sınıf ve branş öğretmenlerinden oluşan 226 kişi araştırmanın çalışma grubunu oluşturmuştur. Araştırma kapsamında geçerlik ve güvenirlik çalışmaları ile psikometrik nitelikleri sorgulanmış olan 38 maddeden oluşan öğretmenlerin tek tip öğrenci kıyafetine ilişkin tutumunu belirlemeyi amaçlayan Tek Tip Öğrenci Kıyafetine İlişkin Öğretmen Tutum Ölçeği kullanılarak veri toplanmıştır. Toplanan veriler üzerinden iki yaklaşım gözetilerek madde ayırt edicilik parametreleri kestirilmiştir. KTK için madde ayırt edicilik parametreleri nokta çift serili korelasyon katsayıları ile POMTK için ise ölçeklenebilirlik katsayıları ile hesaplanmıştır. Veri analizleri R programında yer alan “ctt”, “mokken” paketleri ve SPSS 22.0 kullanılarak gerçekleştirilmiştir. Hesaplanan parametreler arasındaki ilişkiler oldukça yüksek ve manidar bulunmuştur. Bulgulara göre madde ayırt edicilik parametresi kestirimlerinde KTK ve POMTK yaklaşımların birbirinin yerine kullanılabilceği sonucuna ulaşılmıştır.

Anahtar sözcükler: madde ayırt ediciliği, klasik test kuramı, parametrik olmayan madde tepki kuramı

Bu makaleye atf vermek için:

Mor Dirlik, E. (2021). Farklı kuramlar kapsamında hesaplanan madde ayırt edicilik parametrelerinin karşılaştırılması. *Trakya Eğitim Dergisi*, 11(2), 732-744.

Cite this article as:

Mor Dirlik, E. (2021). The comparison of item discrimination parameters estimated from different test theories. *Trakya Journal of Education*, 11(2), 732-744.

ABSTRACT: The main purpose of this study is to compare the item discrimination parameters estimated from Classical Test Theory(CTT) and Nonparametric Item Response Theory(NIRT) and to analyze the similarities of the estimations. For this purpose, a correlational research method included in the quantitative approach was preferred. The study group of this research was composed of 226 class and branch teachers who are in service at primary and elementary public schools. The data of the study was gathered by using the teachers' attitude scale towards students' uniforms composed of 38 items. Item discrimination parameters were calculated using total scores obtained from the scales considering the two test approaches. The item discrimination parameters estimated from CTT by point-biserial correlation coefficients. For NIRT, scalability coefficients were estimated. The data were analyzed via R program with “ctt” and “mokken” packages and SPSS 22.0. The correlations between the parameters were found significant and very high. Considering the findings, while estimating the item discrimination parameters, CTT and NIRT approaches may be used interchangeably.

Keywords: item discrimination parameters, classical test theory, nonparametric item response theory

EXTENDED ABSTRACT

Introduction

Test development is a systematic process that includes several essential steps to be followed by the developers. In order to obtain valid and reliable scores from the developed scales, these steps should be applied carefully and detailed analyses on the total scale scores and items of the scales have to be administered. One of these analyses is the item based analyses and the items' psychometric qualities are investigated in this step. For the items, difficulty and discrimination indexes are calculated. In addition to these indexes, distractor analyses, item variances and reliabilities are calculated. Among these psychometric features of the items, item discrimination parameters are the most essential one and is also named as item

¹ Dr. Öğr. Üyesi, Kastamonu Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri ABD, e-posta: emor@kastamonu.edu.tr, ORCID: 0000-0003-0250-327X

validity. Hence so as to make valid assessments, items should have positive and high item discrimination parameters. Thanks to high discrimination parameters, the ones who have high abilities get higher scores than the ones who have low abilities, so the interpretation of the scale scores becomes easier and meaningful. Also the items are discarded from the scales firstly considering the discrimination parameter values then the difficulty parameters and the other are taken into account. Due to the importance of this parameter for the scale development, there are many ways to estimate this parameter.

These ways are depended on the test theories and one of the best known test theories is Classical Test Theory (CTT). Since the day the theory first developed in the middle of the 1900s, it has been most used theory. Because of the easily met assumptions and simple calculations, it has been preferred so frequently. In the context of CTT, the item discrimination parameters are estimated with several ways, one of these ways is point-biserial correlation coefficients calculated between the items. The coefficients indicate the consistency levels of items and are expected to be as high as possible, but the values are higher than 0,90 are interpreted unacceptable because these items are thought to be exactly same.

The other testing approach is Non-Parametric Item Response Theory (NIRT) and it is more similar to the CTT. It requires less assumptions and these assumptions are met easier than the ones that are required by the precedent of this approach, Parametric Item Response Theory, which requires strict assumptions like large sample size and item pool. However, NIRT models allow scalability in case of less items and small samples. In NIRT, items are also analyzed and the difficulty levels of items are calculated like in CTT as item popularities. The discrimination of items is estimated as scalability coefficients based on item covariances. NIRT provides more accurate psychometric properties than CTT and allows to investigate several different properties like nonchanging item ordering. Due to the advantages over CTT and similarities with that approach, it is important to investigate the suitability and consistency between the estimates of these two approaches. Because NIRT is a newly developed approach relative to CTT, it has many issues to be investigated deeply. Scalability coefficients are one of these issues, so in this study it is aimed to analyze the item scalability coefficients and compare these coefficients to the point-biserial coefficients, which are the indicators of item discriminations in CTT.

Method

This study is designed as correlational method included in qualitative approach. It is not aimed to generalize the findings to the population, so purposive sampling techniques was used. The study group of this research was composed of 226 class and brach teachers who are in service at primary and elementary public schools. All of the participants attended to the study voluntarily. The data of the study was gathered by using the teachers' attitude scale towards students' uniforms composed of 38 items. Item discrimination parameters were calculated using total scores obtained from the scales considering the two test approaches. The item discrimination parameters estimated from CTT by point-biserial correlation coefficients. For NIRT, scalability coefficients were estimated. The data were analyzed via R program with "ctt" and "mokken" packages and SPSS 22.0.

Findings

The item discrimination parameters estimated from CTT were compared with the scalability coefficients estimated from NIRT. The first findings is that both coefficients are found significantly correlated. Secondly, it was found that there are more items which are determined as inappropriate to scaling according to the NIRT approach. In NIRT models, 9 items, in CTT 6 items are defined as lower discriminative power in order to retain in the scale. These 6 items are the same for both approaches. Also the reliability of the scale is investigated by computing four different coefficients and all of the coefficients are found higher than 0,90 which is the critical value for reliability.

Discussion and Conclusion

Considering the findings of the study, it may be accepted that when the item parameters are priority for assessments, NIRT can be used instead of CTT, which provides weaker scale properties. The point biserial correlations and scalability coefficients were found significant and highly correlated. When parametric IRT approach can not be applicable because of small sample size, short test length and strict assumptions, instead of CTT, NIRT models may be preferred which enables to analyze the scale items in an explanatory way and provides more information about the tested constructs.

GİRİŞ

Psikolojik testlerin kullanımı farklı alanlarda gün geçtikçe yaygınlaşmakta, tıptan iletişime, sosyolojiye kadar geniş bir alanda bireyler hakkında bilgi toplamak için psikolojik testlerden yararlanılmaktadır (Özdemir, 2014). Testlerin kullanım amaçları içinde bireyleri tanımak, bir mesleğe yönlendirmek ve bir üst öğrenim kurumu için seçme ve yerleştirme yapmak, bilgi toplamak gibi amaçlar gibi yer almaktadır. Testlerden elde edilen bilgilere göre kararlara varılmakta ve bu kararlardan bazıları, örneğin seçme ve yerleştirme, insanların hayatlarında belirleyici role sahip olmakta ve bireyleri geri dönülmesi oldukça zor olan durumlara yönlendirmektedir. Dolayısıyla bu kararların isabetli olması gerekmekte ve bu durumunda temel olarak kullanılan psikolojik testlerin niteliklerine bağlı olarak değişmektedir. Psikolojik testlere bağlı kalınarak doğru kararlara ulaşılması için ise testlerin sistematik ve bilimsel bir yöntemle bağlı kalınarak geliştirilmesi, gerekli aşamaların takip edilmesi, ilgili psikometrik niteliklerinin sorgulanması ve madde ile ölçek düzeyinde analizlerin yapılması gerekmektedir (Baykul, 2010; Erkuş, 2003).

Test geliştirme çalışmalarında genel amaç az sayıda madde ile yüksek geçerlik ve güvenilirliği sağlayarak amaçlanan alanda ölçme yapabilen bir teste ulaşmaktır. Bu amaca ulaşmak için geniş bir madde havuzunun test edilmesi ve testin geçerliliğine ve güvenilirliğine en fazla katkı yapan maddelerin belirlenerek teste dahil edilmesi gerekmektedir. Yeni test geliştirirken ya da mevcut bir testin daha kısa formunu oluşturmaya çalışırken, madde analizi ve uzman görüşleri temel alınarak hangi maddelerin teste kalacağına karar verilir. Bu nedenle madde analizinden elde edilen veriler, testin bütünü, elde edilen puanların anlamlılığı ve psikometrik nitelikleri yüksek ölçme sonuçlarına ulaşmak için oldukça önemlidir. Madde analizi test alanlarının her bir maddeye verdikleri yanıtlar gözetilerek gerçekleştirilen istatistiksel hesaplamalar ve incelemeleri içeren bir genel bir terimdir (Crocker ve Algina, 2006).

Testlerin temel birimi olan madde parametreleri testin genel niteliğini belirlemekte, dolayısıyla madde analizinden elde edilen veriler testin psikometrik niteliklerini doğrudan etkilemektedir. Bu nedenle testlerden elde edilen puanlar analiz edilirken öncelikle madde analizi ile başlanır (Baykul, 2010). Madde düzeyinde hesaplanabilecek çok sayıda parametre vardır, fakat bunlardan en sık kullanılanları ve ölçüt olarak alınanları madde güçlük ve ayırt edicilik parametreleridir. İki kategorili puanlanan maddeler için madde ortalaması olarak da adlandırılan ve maddenin doğru yanıtlanma düzeyini belirten madde güçlük parametresi ile maddenin ölçülen niteliğe göre alt ve üst grupta yer alan bireyleri ayırma gücünü gösteren madde ayırt edicilik parametresidir.

Birçok testin temel amacı bireysel farklılıklara ilişkin bilgi sağlamaktadır ve bu amaç ancak testte ölçülen niteliğe yüksek düzeyde sahip olanlar ile görece olarak düşük düzeyde sahip olanları ayırabilen maddelerin teste dâhil edilmesi ile gerçekleşir. Örneğin bir başarı testinde bir maddeyi konuyu bilenlerin doğru yanıtlama olasılığının bilmeyenlerden daha yüksek olması beklenir. Eğer testten yüksek puan alanlar ile düşük puan alanların maddeyi doğru yanıtlama olasılığı eşit ise, bu durumda maddeye şüphe ile yaklaşırlar ve maddenin test ile ölçülen yapıdan farklı bir özelliği ölçtüğü düşünülür. Düşük puan alanların çoğunluklu olarak doğru yanıtladığı, yüksek puan alanların ise çoğunlukla yanlış yanıt verdikleri maddeler ise daha istenmeyen bir durumdur ve bu maddelerin negatif ayırt edicilik gösterdikleri belirtilir (Crocker ve Algina, 2006). Özetle ayırt edici bir maddenin, ölçülen özelliğe yüksek düzeyde sahip olanlar tarafından doğru yanıtlanması, düşük düzeyde sahip olanlar tarafından yanlış yanıtlanması ya da yanıtız bırakılması beklenmektedir. Madde geçerliği olarak da bilinen bu istatistik, farklı test kuramları kapsamında farklı şekillerde hesaplanmaktadır.

Klasik Test Kuramı (KTK) ölçme alan yazınında ilk geliştirilen, en fazla yer bulan ve uygulanan kuramlardandır. Bu kuram gerçek puan teorisine dayanmakta ve gözlenen puanı gerçek puan ile hata puanının toplamı olarak ele almaktadır. Hambleton ve Jones (1993)'a göre KTK'nın temel avantajı kuramsal olarak görece zayıf varsayımlar gerektirmesi ve bu sayede çok fazla veri setine uyum sağlamasıdır. KTK'nın test geliştirmede kullanışlı bir model olduğu çok sayıda çalışmada kanıtlanmış bir durumdur ancak bu söz konusu kuramın önemli dezavantajlarını gidermeye yetmemektedir. Kuram kapsamında hesaplanan iki temel parametre; madde güçlük ve ayırt edicilik indeksi, örneklem dağılımına bağımlı olarak hesaplanmaktadır. Bu durumda da bireyin belli bir maddeye ilişkin olası performansını tahmin etmeyi zorlaştırmakta, bireyin yetenek düzeyine ilişkin test oluşturmaya imkân tanımamakta ve değişmeyen parametreler kestirmeye izin vermemektedir (Hambleton, 1991; 1994).

KTK'ya yapılan bu eleştiriler üzerine, 20. yüzyılın sonlarına doğru yeni bir kuram olan Madde Tepki Kuramı (MTK) geliştirilmiş ve bu kuram KTK'nın önemli bir tamamlayıcısı haline gelmiştir. KTK halen yaygın olarak kullanılmaya devam etse de, MTK giderek daha popüler olmaya ve sıkça tercih edilmeye başlanmıştır (Reise, Ainsworth ve Haviland, 2005). Bireyin test performansı ile ölçülen özelliğe ilişkin yeteneğini matematiksel bir fonksiyon üzerinde tanımlayan MTK, bilgisayar teknolojisindeki gelişmeler

üzerine geniş kullanım alanı bulmuş ve teorik olarak da büyük gelişmeler göstermiştir. MTK iki temel varsayım üzerine kurulmuştur. Bunlar;

1. Bireyin bir test maddesindeki performansı yetenek, gizil yetenek ya da yeterlik adı verilen bir takım faktörler tarafından kestirilebilir.
2. Bireyin madde performansı ile bu performansın altında yatan gizil yetenek arasında ilişki madde karakteristik fonksiyonu ya da madde karakteristik eğrisi olarak adlandırılan monoton artan bir fonksiyon ile gösterilir (Hambleton ve Swaminathan, 1985; Embretson ve Reise, 2000).

MTK, KTK'ya göre daha teori temelli bir kuramdır ve bireyin başarısını olasılıksal olarak modellemeye olanak sağlar. Adından da anlaşıldığı üzere, MTK madde yönelimli bir kuramken, KTK daha test yönelimli bir kuramdır. MTK'da bireyin yeteneği maddelerin parametrelerine göre kestirilirken, KTK'da toplam puan kullanılmaktadır. Ayrıca MTK'nın KTK'ya göre temel üstünlüğü madde parametrelerinin bireylerden bağımsız olarak hesaplanmasıdır (Hambleton, 1994). Bu durumda testin ve maddelerin özelliklerinin bireylerin yetenek dağılımından, cinsiyetlerinden, grubun ortalama performanslarından bağımsız olarak hesaplanabilmesine olanak sağlamaktadır. Ayrıca madde parametre değişmezliği de bireye uyarlanmış test geliştirmeyi olanaklı hale getirmektedir. Fakat MTK'nın bu avantajlarının ortaya çıkması için bir takım temel varsayımların karşılanması gerekmektedir. Bunlar; tek boyutluluk, yerel bağımsızlık, testin hız testi olmaması ve model veri uyumudur (Hambleton, Swaminathan ve Rogers, 1991; Linn, 1998). Bahsi geçen varsayımların karşılanamadığı durumlar için farklı MTK uygulamaları geliştirilmiştir.

DeMars (2010) MTK normallik varsayımını gerektirmediğini ancak normal dağılımın sağlanmadığı durumlarda yetenek kestirimlerinde sapmalar ve yanıltıcı kestirimler olabileceği belirtmiştir. Dolayısıyla MTK'nın bahsi geçen avantajlarının elde edilebilmesi için normale uygun dağılımlar gerekmekte ve bu dağılımların sağlamadığı durumlarda ise Parametrik Olmayan Madde Tepki Kuramı (POMTK) yaklaşımlar uygulamaya koyulmaktadır. POMTK yaklaşımı küçük örneklemlerde ve madde sayısının az olduğu testlerde normallikten büyük sapmalar olduğu durumlarda parametrik yaklaşımının alternatifi olarak kullanılmaktadır. Dağılımların çarpık olduğu durumlarda Parametrik Madde Tepki Kuramından elde edilen kestirimler yanıltıcı sonuçlar ortaya çıkarmaktadır (Embretson ve Reise, 2000). POMTK, parametrik yaklaşımlardan farklı olarak madde tepki fonksiyonlarının herhangi bir parametrik fonksiyona bağlı kalınmadan kestirilmesini sağlar, böylelikle PMTK'ya göre daha fazla veri setine ve uygulamalı verilere daha yüksek düzeyde uyum sağlar (Junker ve Sijtsma, 2001).

Daha az varsayım gerektiren ve esnek modelleme sağlayan POMTK kapsamında, Guttman ölçekleme tekniğinin uzantısı olan Mokken Ölçek Analizi (Mokken Scale Analysis) tekniği ile madde ve test analizi yapılmaktadır. Parametrik modellerin doğrulayıcı yaklaşımından farklı olarak açıklayıcı bir yaklaşımla ölçek geliştirme imkânı sunan POMTK, her ne kadar esnek bir modelleme sağlasa da bir takım varsayımların karşılanmasını gerektirmektedir. Bu varsayımlar; tek boyutluluk, yerel bağımsızlık, madde tepki fonksiyonlarının monotonluğu ve kesişmeyen madde tepki fonksiyonlarıdır. İlk üç varsayımın karşılandığı veri setleri Mokken modelleri içinde esnek modellemenin sağlandığı Monoton Homojenlik Modeline (MHM) uyum sağlarken, kesişmeyen MTF varsayımı karşılandığında ise Çift Monotonluk Modeline uyum sağlar (Molenaar ve Sijtsma, 2002; Van Schuur, 2011). Her iki yaklaşım için de karşılanması gereken varsayımlar olmasına rağmen, POMTK, parametrik tekniklerin kullandığı sınırlayıcı madde tepki fonksiyonu kestirim yöntemleri yerine veri yönelimli madde tepki fonksiyonu kestirmeyi olanaklı hale getirmektedir. POMTK'dan elde edilebilecek temel avantaj, madde tepki fonksiyonlarında sınırlama olmadığı için daha gerçekçi tepki fonksiyonları üretebiliyor olmasıdır. Ölçülen nitelik ile madde arasındaki ilişkiyi daha gerçekçi bir şekilde incelemeye olanak sağlamakta, böylece psikolojik yapıları açıklayıcı bir şekilde inceleme olanağı sağlamaktadır. Bunlara ek olarak, parametrik tekniklerin önkoşulu olan normal dağılım varsayımını gerektirmiyor olması, daha küçük örneklemler ve daha az madde ile ölçekleme yapmayı olanaklı hale getirmektedir (Molenaar and Sijtsma, 2002).

POMTK kapsamında ölçek ve madde düzeyinde analizler yapılmaktadır. Ölçek düzeyinde bireylerin toplam puanları kullanılırken, madde düzeyindeki analizler PMTK'ya göre daha zayıf kalmaktadır. Madde güçlük indeksi KTK'daki madde güçlük indeksi ile aynı şekilde hesaplanır ve yorumlanır. Madde ayırt edicilik indeksi için ise Loevinger (1947, 1948) tarafından önerilen ölçeklenebilirlik katsayısı kullanılmaktadır (van Onna, 2004). H ile gösterilen ölçeklenebilirlik katsayısı, Guttman ölçekleme tekniğinde yer alan Guttman hatası kavramıyla benzerdir. Guttman hatası, bireyin nispeten daha zor bir maddeye doğru yanıt verirken daha kolay bir maddeye yanlış yanıt vermesiyle ortaya çıkmaktadır. Ölçeklenebilirlik katsayısı bireyleri ölçülen özelliğe göre sıralamada maddenin ne kadar

uyumlu olduğunu gösterir (akt. Kuijpers, van der Ark, Croon, 2013; Sijtsma ve Molenaar, 2002). Ölçeklenebilirlik katsayıları madde çiftleri için H_{ij} , maddeler için H_i ve tüm ölçek için H olmak üzere üç farklı şekilde hesaplanmaktadır. Maddeler arası kovaryansa dayalı olarak hesaplanan ölçeklenebilirlik katsayıları toplam puanın varyansı ile artan bir fonksiyon göstermektedir (Mokken, 1971). MHM kapsamında yüksek ölçeklenebilirlik katsayısı değerleri yüksek ayırt edicilik göstergesi olarak kabul edilmektedir. Yüksek ölçeklenebilirlik katsayısına sahip maddelerden oluşan ölçekten elde edilen toplam puana göre birey sıralaması daha güvenilir olmaktadır (De Koning, Sijtsma, & Hamers, 2002). Ayrıca madde ölçeklenebilirlik katsayısı yüksek olan maddeler grup içindeki bireyleri performanslarına göre iyi ayırabilmektedir. Bu nedenle Meijer ve Baneke (2004) madde ölçeklenebilirlik katsayılarının PMTK yaklaşımındaki a parametresinin yerine kullanılabilmesini belirtmiştir. Uygulamada madde ve ölçek düzeyindeki H katsayıları 0 ile 1 arasında değişmekte ve 0'a yakın H_i değerleri yatay MTF'leri gösterirken, 1'e yakın değerler ise dik MTF'leri göstermektedir. Mokken (1971) ölçeklenebilirlik katsayıları için sınır değerlerin 0.3 olarak alınması gerektiğini, 0.3 ile 0.4 arasındaki değerlerin düşük, 0.4 ile 0.5 arasındaki değerlerin ise yüksek düzeyde ölçeklenebilirliği gösterdiğini belirtmektedir.

İlgili çalışmalar incelendiğinde, POMTK yaklaşımını PMTK modellerine tamamen alternatif olması değil, gerekli durumlarda tüm psikometrik yaklaşımlarla beraber kullanılması gerektiği vurgulanmaktadır. Buna ek olarak POMTK modelleri kapsamında veri yapısına dayalı olarak kestirilen MTF'lerin ölçülen psikolojik yapıların daha iyi anlaşılmasını, keşfedilmesini sağlayacak nitelikte olduğu belirtilmektedir. Özellikle uygulamada oluşabilecek farklı durumların incelenmesi alan yazına katkıda bulunacağı ifade edilmektedir (Christensen ve Kreiner, 2012; Junker ve Sijtsma, 2001; Van Schuur, 2011). Bu bilgiler gözetilerek alan yazında yer alan madde parametrelerini karşılaştıran çalışmalar incelenmiştir. Fan (1998) gerçek verilere dayalı olarak yaptığı çalışmada KTK ve PMTK'nın madde ve yetenek parametreleri arasındaki korelasyonları incelemiş ve her iki yaklaşımdan kestirilen yetenek ve madde parametreleri arasındaki ilişkiyi 0.90'ın üzerinde ve manidar bulunmuştur. Courville (2005) tarafından yapılan çalışma kapsamında da Fan'ın çalışmasından elde edilen bulgulara oldukça benzer bulgular bulunmuş ve madde ve yetenek parametreleri arasında yüksek ve manidar korelasyonlar olduğu belirlenmiştir. McDonald ve Paunonen (2002)'in simülatif verilere dayanan çalışmalarında, KTK ve PMTK'dan kestirilen madde güçlük parametreleri arasında yüksek ve manidar bir ilişkinin olduğunu bulmuşlardır. Ayrıca her iki yaklaşımdan kestirilen madde ayırt edicilik parametrelerinin, ancak madde güçlük indekslerine ilişkin varyansın küçük, ayırt edicilik indekslerine ilişkin varyansının da büyük olduğu durumlarda yüksek korelasyon gösterdiği bulunmuştur. Progar ve Socan (2008) yaptıkları çalışmada ise gerçek veri seti üzerinden kestirilmeye gitmiş ve MTK ve KTK'dan elde ettikleri madde ve yetenek parametrelerini karşılaştırmışlardır. Madde parametrelerinin değişmezlik özelliklerinin incelendiği bu çalışmada, her iki yaklaşımında benzer düzeyde değişmezliği sağladığını ve parametreler arasındaki korelasyonun da yüksek ve manidar olduğu bulunmuştur.

Kuramsal olarak yapılan karşılaştırma çalışmaları yerli alanyazında da mevcuttur. Özellikle geniş ölçekli test uygulamalarının verilerinin kullanıldığı ve bu maddelere ilişkin parametre karşılaştırmalarının yapıldığı çalışmaların (Kılıç, 1999; Yalçın, 1999; Çalışkan, 2000; Çelik, 2001; Karataş, 2001; Özkurt, 2002; Can, 2003; Yapar, 2003; Yeğin, 2003) sayısı fazladır. Öğretmen yapımı testlere ilişkin yapılan karşılaştırma çalışması da, her iki kuramdan benzer kestirilmeye ulaşıldığını belirlemiştir (Çelen ve Aybek, 2013). Ancak bu çalışmaların tümü KTK ile PMTK'dan kestirilen parametreleri kıyaslamayı amaçlamıştır ve POMTK ve KTK'dan kestirilen parametreleri karşılaştırmayı amaçlayan çalışmalara alan yazında rastlanamamıştır. İki yaklaşım esnek varsayımlarından dolayı benzerlik göstermekte ve PMTK'ya göre daha fazla veri setine uyum sağlamaktadır. POMTK, KTK'dan farklı olarak modelleme imkânı sağlamakta ve açıklayıcı bir yaklaşımla maddelere ve ölçülen yapıya ilişkin keşfedici bir bakış açısı sağlamaktadır (Molenaar, 2001). Hem KTK gibi küçük örneklem ve az sayıda madde ile kestirim yapmaya olanak sağlaması hem de özellikle alanyazında yeni olması ve araştırmacılar tarafından yeterince tanınmadığından KTK'ya alternatif olarak görülmemesinden dolayı, bu çalışma kapsamında madde parametrelerinin her iki kuramdan da kestirilmesi ve incelenmesi gerekli görülmüştür.

Çalışma kapsamında farklı iki kuramdan kestirilen madde ayırt edicilik parametrelerinin karşılaştırılması ve benzerliklerinin incelenmesi amaçlanmıştır. Bu indeksin seçilme nedeni ise madde ayırt edicilik gücünün belirlenmesinde farklı psikometrik kuramlar ve yaklaşımlar tarafından çok sayıda istatistik önerilmiş olması ve bunları kıyaslayan çalışmaların az sayıda olmasıdır. Psikometri alanında en çok tercih edilen ve ilk geliştirilen yaklaşım olan KTK kapsamında, nokta-çift serili (point-biserial), çift-serili (biserial), dörtlü(phi) ve tetrakorik (tetrachoric) korelasyon katsayıları, alt-üst grup farkına dayalı katsayılar olmak üzere birden çok teknikte ayırt edicilik değeri hesaplanabilmektedir. POMTK kapsamında ise daha önce de belirtildiği gibi Loehinger tarafından geliştirilen ölçeklenebilirlik katsayıları kullanılmakta ve

madde ayırt edicilik indeksi olarak yorumlanmaktadır. Maddeler arasındaki koşullu kovaryanslar temel alınarak hesaplanan ölçeklenebilirlik katsayılarının yalnızca POMTK modelleri ile sınırlı kalmaması gerektiği, maddeler hakkında bilgi toplamak için tüm yaklaşımlarda bu katsayılardan yararlanılabileceği farklı araştırmacılar tarafından önerilmektedir (Junker ve Molenaar, 2001; Sijtsma ve Molenaar, 2002; Meijer ve Baneke, 2002; Van Schuur, 2011). Farklı ölçekleme yaklaşımları benimsenerek geliştirilmiş ölçeklerdeki maddelere ilişkin ölçeklenebilirlik katsayılarının hesaplanması maddeler için daha ayrıntılı analizler yapılmasını sağlayacak, test geliştiricilere ölçülen yapı hakkında daha ayrıntılı bilgiler sağlayacak ve maddeleri ve ölçülen yapıları keşfetmeye katkıda bulunacaktır. İlgili çalışmalar incelendiğinde Mokken modellerinin KTK ve MTK'ya göre daha yakın zamanda psikometri alanına dahil olmasından dolayı farklı yaklaşımları karşılaştırmayı amaçlayan çalışmalarının sayısının az olduğu sonucuna ulaşılmaktadır.

Araştırmanın Amacı

Çalışmada KTK temel alınarak geliştirilmiş, çok kategorili olarak puanlanan ve öğretmenlerin tek tip öğrenci kıyafetine ilişkin tutumlarını belirlemeyi amaçlayan bir tutum ölçeğindeki maddelere ilişkin madde ayırt edicilik düzeylerinin KTK ve POMTK yöntemleri ile kestirilmesi ve kestirimler arasındaki korelasyon katsayılarının belirlenmesi amaçlanmıştır. Ölçeğin KTK'ya göre geliştirilmiş olması POMTK ile ölçekleme yapmaya engel oluşturmamakta, aksine ölçeğin psikometrik niteliklerinin derinlemesine incelenmesini sağlamaktadır. Bu genel amaca göre yanıt aranan alt araştırma soruları şu şekildedir:

1. Tek Tip Öğrenci Kıyafetine İlişkin Öğretmen Tutum Ölçeği Mokken modellerinden hangisine uyum göstermektedir?
2. Ölçeğin uygun olduğu Mokken modeline göre kestirilen madde düzeyindeki ölçeklenebilirlik katsayıları nasıldır?
3. Ölçeğin KTK'ya göre hesaplanan madde ayırt edicilik indeksleri nasıldır?
4. Her iki yaklaşımdan kestirilen madde ayırt edicilik parametreleri arasında manidar bir ilişki var mıdır?
5. Ölçümlerin güvenilirliği iki kurama göre nasıl değişmektedir?

YÖNTEM

Araştırmanın Modeli

Çalışma kapsamında farklı kuram ve yaklaşımlar çerçevesinde kestirilen madde ayırt edicilik parametrelerinin karşılaştırılması amaçlanmıştır. Bu genel amaç gözetildiğinde çalışmanın nicel araştırmalar içinde yer alan betimsel bir çalışma olduğu ve korelasyonel araştırma deseniyle yürütülmesi uygun bulunmuştur.

Çalışma Grubu/ Evren- Örneklem

Çalışma kapsamında seçkisiz olmayan örneklem belirleme yöntemlerinden uygun örnekleme kullanılmıştır. Uygun örnekleme yöntemi, araştırmacının evrenden seçkisiz ya da sistematik bir şekilde örneklem belirlemeye gitmediği durumlarda, mevcut gruplarla yürütülen çalışmalar için tercih edilen bir yöntemdir (Fraenkel ve Wallen, 2006). Veriler ilkökul ve ortaokullarda görev yapan 226 öğretmenden toplanmıştır. Katılımcıların 110'u erkek, 116'sı ise kadındır. Yaşları ise 22 ile 53 arasında değişmektedir. Görev yapılan okul türüne göre incelendiğinde ise, 116 katılımcı ilkökulda, 110 katılımcı ise ortaokulda çalışmaktadır.

Veri Toplama Aracı

Veri toplama aracı olarak ise araştırmacı tarafından önceki çalışmaları kapsamında geliştirilen "Tek Tip Öğrenci Kıyafetine İlişkin Öğretmen Tutum Ölçeği" kullanılmıştır. Bu ölçeğin seçilme nedeni tek boyutlu bir yapıda olması ve test puanlarının geçerlik ve güvenilirlik gibi psikometrik niteliklerine ilişkin çalışmaların yapılmış olmasıdır. Likert tipi beş kategorili olarak puanlanan 39 maddeden oluşan tutum ölçeği geliştirilme çalışmasında öncelikli olarak hedef kitle olan bir grup ilk ve ortaokul öğretmenlerine açık uçlu birkaç maddeden oluşan bir bilgi alma formu uygulanmış ve öğretmenlerin genel eğilimleri belirlenmeye çalışılmıştır. Öğretmenlerin açık uçlu maddelere verdikleri yanıtlar ve alan yazındaki benzer çalışmalar incelenerek denemelik tutum ölçeği formu oluşturulmuş ve uzman görüşlerine sunulmuştur. Uzmanların görüşlerine göre revize edilen tutum ölçeği 42 maddelik deneme formu olarak hazırlanarak, 232 öğretmene uygulanmıştır. Elde edilen verilerle ölçme sonuçlarının geçerlik ve güvenilirlik nitelikleri incelenmiştir. Geçerlik çalışmaları kapsamında kapsama geçerliği için uzman kanılarına başvurulmuş, yapı

geçerliđi için açılmayıcı ve dođrulamayı faktör analizleri yapılmıřtır. İ ölçüte dayalı geçerlik için ise alt-üst grup ortalamaları farkına dayalı t-testi ile test edilmiřtir. Güvenirlik alıřmalarında ise iç tutarlılık ve kararlılık incelemeleri yapılmıřtır. İ tutarlılık anlamında güvenirlik için Cronbach alfa katsayısı hesaplanmış ve 0,90 olarak bulunmuřtur. Test-tekrar test uygulamasında ise 0,88 düzeyinde güvenirlik katsayısına ulařılmıřtır. Bu alıřma kapsamında kullanılan ölçeđe iliřkin geçerlik incelemesi için dođrulamayı faktör analizi yapılmıřtır. Tek faktörlü modele iliřkin elde edilen istatistikler incelendiđinde, modelin manidar olduđu ($p \leq 0,05$), tüm maddelerin t deđerlerinin de ($p \leq 0,05$) manida olduđu bulunmuřtur. Model-veri uyumuna iliřkin istatistikler incelendiđinde (GFA=0,94; AGFI=0,90; CFI=0,92; IFI=0,94; RMSEA= 0,005, $\chi^2/sd=2,75$), genel düzeyde kabul edilebilir ve iyi düzeyde uyumun olduđu bulunmuřtur. Ölçümlerin güvenirliđi için ise Cronbach alfa iç tutarlılık katsayısı hesaplanmış ve 0,92 olarak bulunmuřtur. Özetle geliřtirilme ve geçerleme alıřmaları ayrıntılı ve uygun bir řekilde yapılan ve öđretmenlerin öđrenci kıyafetlerine iliřkin tutumlarını belirlemeyi amalayan bu ölçek alıřma kapsamında madde parametrelerinin incelenmesi için veri toplama aracı olarak kullanılmıřtır.

Verilerin Toplanması ve Analizi

Verilerin toplanması öncesinde katılımcılara arařtırmanın amacı açıklanmış ve alıřma için bireyler motive edilmiřtir. Katılımcılar gönüllülük esasına dayalı olarak alıřmaya dahil olmuşlar ve veriler ders dıřı zamanlarda toplanmıřtır.

Belirtilen alıřma grubundan elde edilen veriler alıřmanın temel amacı gözetilerek POMTK ve KTK'ya dayalı analizler yapılmıřtır. Herhangi bir analiz yapılmadan önce kayıp veriler incelenmiřtir. EM (expectation-maximization) algoritması kullanılarak yapılan kayıp veri incelemesinde, bir maddeye iliřkin kayıp deđere sahip olan katılımcı sayısı %3, iki maddeye iliřkin kayıp deđere sahip olan katılımcı sayısı ise %4 olarak bulunmuřtur. Kayıp veri oranı %5'in altında olduđu için (Field, 2005), herhangi bir dönüřtürme veya atama yoluna gidilmemiř, kayıp veriye sahip katılımcıların yanıt öđrüntüleri veri seti dıřında bırakılmıřtır.

Analiz ařamasına öncelikli olarak yaklařımların varsayımlarının incelenmesi ile bařlanmışır. POMTK yaklařımı için tek boyutluluk, yerel bađımsızlık, monotonluk ve deđiřmez madde sıralaması varsayımları test edilmiřtir. KTK için ise normallik ve boyutluluk incelemesi yapılmıřtır. Mokken modellerine göre tek boyutluluk incelemesi için Otomatik Madde Seçim İřlemi (OMSİ) tekniđi kullanılmıřtır ve sınır deđer olarak 0,3 kabul edilmiřtir (Mokken, 1971; Sijtsma ve Molenaar, 2002). Monotonluk incelemesi için Madde Tepki Fonksiyonları kestirilmiş ve incelenmiřtir. Yerel bađımsızlık varsayımı ise artık korelasyon matrisi oluřturularak incelenmiřtir Deđiřmez madde sıralaması varsayımı ise P matrisi, madde kalan puanı grubu ve H^T katsayısı incelenmiřtir. Deđiřmez madde sıralaması varsayımı ihlalinin deđerlendirilmesinde madde düzeyinde sınır deđer 80 olarak alınmıřtır. Madde adım faonksiyonlarında 80 ve üzerinde kesiřme gösteren maddeler varsayımı ciddi düzeyde ihlal eden maddeler olarak kabul edilmekte ve bu maddelerin için ift Monotonluk Modelinin uygun olmadıđı sonucuna ulařılmaktadır(Sijtsma ve van der Ark, 2017). KTK kapsamında ise normallik incelemesi için betimsel istatistikler ve normallik testleri yapılmıřtır.

Madde analizleri ařamasında madde onaylanma oranları için her iki yaklařımda da madde ortalamalarını belirten p deđerleri hesaplanmışır. Ayırt edicilik indeksleri için KTK'ya dayalı olarak madde düzeltilmiş madde toplam korelasyon katsayıları, POMTK'ya dayalı olarak ise madde ölçeklenebilirlik katsayıları hesaplanmışır. POMTK'ya göre yapılan analizlerde R programı içinde yer alan "mokken" paketi (van der Ark, Koopman, Straat ve van den Bergh, 2020), KTK'ya dayalı olarak yapılan analizde ise SPSS 22.0 kullanılmıřtır. Kestirilen parametrelerin öncelikli olarak betimsel istatistikleri hesaplanmış, ardından kestirilen parametreler arasındaki iliřki Spearman Brown Korelasyon Katsayısı ile incelenmiřtir.

BULGULAR

alıřma kapsamında elde edilen bulgular KTK ve POMTK yaklařımlarına göre ayrı olarak verilmiş ve arařtırma sorularının sıralaması izlenmiřtir.

Mokken Modelinin Belirlenmesi

alıřma kapsamında toplanan veriler Mokken ölçekleme tekniđine göre analiz edilmiřtir. Analizin ilk adımında ölçeklenebilirlik katsayıları hesaplanmışır. 38 maddelik ölçeđe yapılan ilk analizde üç

maddeye ilişkin H_i değerinin 0 olduğu, altı maddenin ise ölçeğin orijinal yapısından farklı boyutlarda ölçeklenebileceği belirlenmiştir. Madde çiftlerine dayalı ölçeklenebilirlik katsayıları incelendiğinde negatif bir değer olmadığı belirlenmiştir. Ayrıca ölçek düzeyindeki H değerinin 0,310 olduğu, dolayısıyla ölçeğin Mokken ölçeklemeye zayıf düzeyde uyum sağladığı belirlenmiştir.

H_i değerleri 0 olarak hesaplanan ve tek boyutta ölçeklenmeye uygun olmadığı belirlenen maddeler sırayla analizden çıkarılmıştır. 29 maddeden oluşan ölçek için ölçeklenebilirlik katsayıları tekrar hesaplanmıştır. Madde çiftleri için hesaplanan ölçeklenebilirlik katsayılarında negatif bir değer olmadığı belirlenmiştir. Yeniden hesaplanan öadde düzeyindeki ölçeklenebilirlik katsayıları ve madde ortalamaları Tablo 1’de verilmiştir.

Tablo 1.

MHM’den Kestirilen Madde Ayırt Edicilik İndeksleri

Maddeler	H_i	Madde ortalaması
M1	0,500	3,81
M2	0,533	3,93
M4	0,532	3,55
M5	0,503	4,24
M6	0,548	3,84
M7	0,527	3,26
M8	0,515	3,61
M11	0,516	3,93
M12	0,317	3,62
M14	0,432	3,81
M15	0,537	3,59
M16	0,311	3,75
M17	0,495	3,06
M19	0,532	3,71
M21	0,465	4,34
M22	0,445	3,77
M23	0,484	3,39
M24	0,469	3,82
M25	0,504	3,37
M27	0,569	3,57
M28	0,549	3,76
M30	0,426	4,13
M32	0,513	3,56
M33	0,503	3,26
M34	0,572	3,84
M35	0,355	4,07
M36	0,489	3,53
M37	0,424	2,96
M38	0,542	3,84
H katsayısı	0.490	

Ölçeklenebilirlik katsayıları incelendiğinde en düşük değer 0,311 olarak 16 nolu maddeye ait olduğu, en yüksek değer ise 34 nolu maddeye ait olduğu 0,572 olarak hesaplandığı belirlenmiştir. Ölçek düzeyindeki H değerinin ise 0,490 olduğu bulunmuş ve bu değer ölçeğin Mokken modellerine orta düzeyde uyum sağladığı anlamına gelmektedir. Madde ortalamaları incelendiğinde ise birçok maddenin 3’ün üzerinde değerler aldığı belirlenmiştir.

Ölçeklenebilirlik katsayılarının ardından monotonluk incelemesi yapılmış ve kritik değer olarak 80 alınmıştır. Monotonluktan madde düzeyinde ihlaller olduğu ancak bu değerlerin kritik değer olan 80’in altında olduğu belirlenmiştir. Dolayısıyla monotonluk varsayımının 29 maddelik ölçek formu için karşılandığı belirlenmiştir. Bu bulgulara göre ölçeğin Monoton Homojenlik Modeline uygun olduğu sonucuna ulaşılmıştır. Mokken modelleri kapsamında daha sınırlayıcı bir model olan Çift Monotonluk Modeline uyumunu incelemek için değişmez madde sıralaması varsayımı incelenmiştir. Bu inceleme için P matrisi ve madde-kalan puanı regresyonu yöntemi kullanılmıştır. Her iki yöntemde de 14 madde düzeyindeki ihlallerin eşik değer olan 80’den yüksek olduğu, 15 maddenin ise değişmez madde sıralaması özelliğine sahip olduğu belirlenmiştir. Bu özelliğe sahip maddelerin numarası şu şekildedir: 21, 5, 30, 35, 11, 8, 27, 32, 4, 36, 23, 25, 33, 7, 17 ve 37’dir. Bu maddelere ilişkin analizler tekrarlanmış ve değişmezliğin netliği hakkında bilgi veren H^T katsayısı hesaplanmıştır. Ancak bu değer sınır değer olan 0,3’ün altında olduğu bulunmuş, dolayısıyla bu ölçeğinde maddeler arasındaki hiyerarşiyi sağlamadığı sonucuna

ulaşmıştır. Sonuçta tutum ölçeğinin 29 maddelik formunun Monoton Homojenlik Modeline uygun şekilde ölçeklenebildiğine karar verilmiş ve KTK kapsamındaki analizlere geçilmiştir.

KTK için madde analizlerine geçilmeden önce veri setinin normalliği incelenmiştir. Madde düzeyindeki incelemelerde tüm maddelerin çarpıklık (0,685, -0,582) ve basıklık değerlerinin -1 ile 1 arasında olduğu belirlenmiştir. Histogram ile yapılan değerlendirmelerde de maddelerin normal dağılıma uygun olduğu sonucuna ulaşılmıştır. Ölçekten elde edilen toplam puana ilişkin yapılan analizler sonucu elde edilen bulgular Tablo 2’de verilmiştir.

Tablo 2.

Ölçek Toplam Puanına İlişkin Betimsel İstatistikler

N	228
Ortalama	133,7675
Ortanca	138,0000
Tepe Değer	144,00
Çarpıklık	-,839
Basıklık	,539
En Düşük Puan	38,00
En Yüksek Puan	182,00

Ölçek toplam puanına ilişkin hesaplanan betimsel istatistikler incelendiğinde, ortalama, ortanca ve tepe değerinin birbirine yakın olduğu görülmektedir. Ayrıca çarpıklık ve basıklık değerlerinin de -1 ile 1 arasında olduğu görülmektedir. Dolayısıyla katılımcıların ölçekten aldıkları puanların dağılımının normal dağılıma uygun olduğu sonucuna ulaşılmaktadır. Ölçekten en düşük 38, en yüksek 182 puanın alındığı, ranjinin de 144 puan olduğu belirlenmiştir. Madde sayısı gözetenilerek alınabilecek en düşük puan 38 en yüksek puanın da 182 olduğu, dolayısıyla olası tüm ranjin kapsandığı sonucuna ulaşılmıştır. Betimsel istatistiklerin ardından madde düzeyindeki analizlere geçilmiş, madde ortalamaları ve madde ayırt edicilik indeksleri hesaplanmıştır. Madde güçlük indeksleri Tablo 1’de verilen ve Mokken ölçekleme ile kestirilen doğru yanıtlanma oranları ile aynı olan p değerleridir, dolayısıyla burada tekrar verilmemiştir. KTK’ya göre hesaplanan düzeltilmiş madde-toplam korelasyon katsayıları ile Monoton Homojenlik Modelinden kestirilen ayırt edicilik değerleri Tablo 3’te verilmiştir.

Tablo 3.

POMTK ve KTK’den Kestirilen Madde Ayırt Edicilik Parametreleri

Kullanılan Yaklaşım	POMTK	KTK	Kullanılan Yaklaşım	POMTK	KTK
Maddeler	H _i	r	Maddeler	H _i	r
M1	0,500	0,684	M20	-----	-0,143
M2	0,533	0,723	M21	0,465	0,569
M3	-----	0,220	M22	0,445	0,600
M4	0,532	0,733	M23	0,484	0,641
M5	0,503	0,640	M24	0,469	0,617
M6	0,548	0,738	M25	0,504	0,677
M7	0,527	0,708	M26	-----	0,271
M8	0,515	0,691	M27	0,569	0,779
M9	-----	0,271	M28	0,549	0,739
M10	-----	-0,472	M29	-----	0,103
M11	0,516	0,699	M30	0,426	0,527
M12	0,317	0,409	M31	-----	0,381
M13	-----	0,344	M32	0,513	0,692
M14	0,432	0,579	M33	0,503	0,665
M15	0,537	0,733	M34	0,572	0,780
M16	0,311	0,414	M35	0,355	0,461
M17	0,495	0,639	M36	0,489	0,669
M18	-----	0,387	M37	0,424	0,553
M19	0,532	0,713	M38	0,542	0,738

Tablo 3’te yer alan istatistikler incelendiğinde, her iki kuram kapsamında kestirilen madde ayırt edicilik değerlerinin benzer eğilimlerde olduğu görülmektedir. POMTK kapsamında ölçeklenebilirlik katsayıları sınır değer olan ,30’dan düşük olan maddelerin çoğunluğunun KTK kapsamında da düşük ayırt ediciliğe sahip maddeler olduğu belirlenmiştir. Benzerliklerin daha ayrıntılı incelenmesi için iki parameter seti arasındaki korelasyonlar incelenmiş, ancak normallik varsayımı sağlanamadığı için parametrik

olmayan yöntemlerden olan Spearman Brown Sıra Farkları Korelasyon Katsayısı hesaplanmıştır. Elde edilen değerler Tablo 4 'te verilmiştir.

Tablo 4.

POMTK ve KTK'ya Göre Kestirilen Madde Ayırt Edicilik Parametreleri Arasında Hesaplanan Spearman Brown Sıra Farkları Korelasyon Katsayısı

	POMTK	KTK
POMTK	1,000	0,987**
KTK	0,987**	1,000

**p<0,01

Tablo 4.'te yer alan korelasyon katsayısı değerinden de görülebileceği gibi, POMTK ve KTK'dan kestirilen madde parametreleri arasındaki ilişki pozitif, çok yüksek ve manidardır. Bu korelasyon katsayısına dayalı olarak POMTK kapsamında kestirilen madde düzeyindeki ölçeklenebilirlik katsayıları KTK'dan kestirilen madde çift serili korelasyon katsayıları ile oldukça uyumlu sonuçlar vermektedir.

Çalışma kapsamında yanıt aranan son soru ölçüm güvenilirliğinin her iki yaklaşıma göre nasıl değiştiğidir. Güvenirlik incelemesine geçilmeden önce, her iki kurama için madde analizi sonuçlarına dayalı olarak, ayırt edicilik değeri düşük bulunan maddeler test dışında bırakılmıştır. POMTK kapsamında MS ve LCRC (Latent Class Reliability Coefficient) katsayısı, KTK kapsamında ise Cronbach Alfa katsayısı ve iki yarı güvenilirliği incelenmiştir. Hesaplanan katsayılar Tablo 5'te verilmiştir.

Tablo 5.

Güvenirlik Katsayıları

POMTK		KTK	
MS Katsayısı	0,960	Cronbach Alfa Katsayısı	0,901
LCRC Katsayısı	0,960	İki Yarım Güvenirliği	0,900
Madde sayısı	23	Madde sayısı	34

Tablo 5'te yer alan değerler incelendiğinde, dört katsayının da sınır değer olan .70'den büyük olduğu görülmektedir. Bu durumda test puanlarının güvenilirliğin yüksek, puanlara karışan hata miktarının az olduğu anlamına gelmektedir. POMTK'dan elde edilen katsayılar ile KTK'dan elde edilen katsayılar oldukça benzer olmasına rağmen, POMTK'dan elde edilen katsayıların ölçeğin 23 maddelik formundan elde edilmiştir. KTK kapsamındaki güvenilirlik incelemesinden elde edilen katsayılar ise ayırt edicilik indeksi düşük olan 4 madde ölçek dılında tutularak toplam 34 madde üzerinden 38 hesaplanmıştır. Buna göre POMTK'ya göre ölçeklenen ve daha az maddeden oluşan ölçeğin güvenilirliği daha fazla maddeden oluşan ve KTK'ya göre ölçeklenen ölçeğin güvenilirliğinin oldukça benzer düzeyde olduğu sonucuna ulaşılabilir. Bu bulguya göre POMTK ile ölçeklerin kısa formlarını güvenilirlikten ödün vermeden oluşturmak olarak mümkün olmaktadır (Molenaar, 2001; Sijtsma, Meijer ve van der Ark, 2011). POMTK ile güvenilirlikten taviz vermeden daha kısa ölçekler geliştirmek mümkündür ve bu durum bu çalışma kapsamında da örneklenmiştir.

TARTIŞMA, SONUÇ ve ÖNERİLER

Bu çalışma kapsamında tipik performansı ölçmeyi amaçlayan bir tutum ölçeğinin madde ayırt edicilik parametre kestirimlerinin iki ayrı ölçekleme yaklaşımı olan KTK ve POMTK'ya göre hesaplanması ve ulaşılan değerlerin karşılaştırılması amaçlanmıştır. Bu amaç kapsamında her iki yaklaşım kapsamında elde edilen madde ayırt edicilik değerlerinin benzer olduğu bulunmuştur. İki yaklaşımda da ölçeklemeye uygun olmadığı belirlenen maddeler aynı maddelerdir ve buna ek olarak madde ayırt edicilik değerleri arasındaki ilişkinin de yüksek ve manidar olduğu belirlenmiştir. Bu bulgu iki ayrı yaklaşıma göre hesaplanan madde ayırt edicilik indekslerinin birbirlerinin yerine kullanılacakları şeklinde yorumlanmıştır.

Mokken modellerine göre dokuz madde ölçeklemeye uygun nitelikte olmadığı için ölçek dışında bırakılmışken, KTK'ya göre altı maddenin sınır değerinin altında ayırt edicilik indeksine sahip olduğu belirlenmiştir. KTK'da belirlenen 6 maddenin Mokken ölçekleme tekniğinde ölçek dışında bırakılan maddeler olduğu bulunmuştur. Dolayısıyla her iki yaklaşıma göre de testten çıkarılması gereken maddeler benzerdir. Bu durumda bir psikometrik kurama göre test dışında kalması gereken bir maddenin farklı bir kurama göre testte yer almasının düşük bir olasılık olduğunu destekler niteliktedir (Çelen, 2008).

Farklı yaklaşımlardan elde edilen madde parametreleri arasındaki ilişkinin manidar ve yüksek bulunması, Berberoğlu (1989)'nun ve Stage (1998)'in KTK ve MTK'dan yaptıkları madde parametre

kestirimlerinin benzer olduğu bulguları ile paralellik göstermektedir. Ayrıca POMTK ile PMTK'dan yapılan madde ve yetenek parametrelerinin de benzerlik gösterdiği Mor Dirlik (2017) ve Koğar (2015)'in çalışmalarıyla uyumludur.

KTK ve POMTK'ya göre incelenen ölçeğin güvenilirlik incelemesi sonucunda, ölçeğin her iki formundan elde edilen puanların benzer ve yüksek düzeyde güvenilirliğe sahip olduğu belirlenmiştir. POMTK ile ölçeklenen ölçek daha az maddeye sahip olmasına rağmen, daha fazla maddeye sahip olan ve KTK ile ölçeklenen form ile benzer düzeyde güvenilirlidir. Dolayısıyla madde sayısının azalması POMTK ile ölçekleme yapıldığı durumlarda güvenilirlikte azalmaya neden olmamaktadır. Bu bulgudan hareketle POMTK'nın ölçeklerin kısa formlarını oluşturmada alanyazında belirtildiği üzere kullanışlı bir yaklaşım olduğu sonucuna ulaşılmaktadır.

Bu çalışma kapsamında kullanılan örneklem büyüklüğü PMTK'ya göre küçük örneklem büyüklüğündedir. KTK'ya göre örneklem sınırlılığı bulunmasa da veri setinin normallik varsayımını karşılaması ve örneklem büyüklüğü arttıkça da normal dağılıma yaklaştığı bilinmektedir. Straat (2012) simülatif verilere dayalı olarak yaptığı ve POMTK'da örneklem büyüklüğünün etkisini incelemeyi amaçladığı çalışmada, örneklem büyüklüğünün yeterliği ile ölçeklenebilirlik katsayılarının ilişkili olduğunu bulmuştur. Bu çalışmada kapsamındaki ölçeklenebilirlik katsayılarının yüksek olması ($H=0,49$) 50 ile 250 kişiden oluşan verinin yeterli olduğu anlamına gelmektedir. Dolayısıyla bu bulgudan hareketle, normallik varsayımın sağlanmadığı küçük örneklemelerde dahi, POMTK kapsamında ölçekleme yapmak ve tek boyutlu ölçekler oluşturmak olanaklıdır.

Bu sonuçlar doğrultusunda, KTK kullanımının zorunlu olduğu, küçük örneklem ya da kısa test gibi durumlarda, KTK'ya alternatif olarak daha güçlü modelleme sağlayan, açıcı bir yaklaşım öneren Mokken modellerinin tercih edilebileceği sonucuna ulaşılmıştır. KTK'ya benzer şekilde esnek ve kolay karşılanan varsayımlara sahip POMTK yaklaşımını kullanılarak ölçülen psikolojik yapıların madde tepki fonksiyonları oluşturularak daha detaylı bir şekilde anlaşılması ve keşfedilmesi sağlanabilir. Ayrıca KTK kapsamında ölçekleme yapılacak olsa dahi madde düzeyinde daha fazla bilgi edinmek için POMTK'nın keşfedici analizlerinden yararlanılabilir.

Bu çalışma KTK ve POMTK kapsamında kestirilen ayırt edicilik indeksleri arasındaki ilişkinin incelenmesi ve her iki yaklaşıma göre oluşturulan ölçeklerin güvenilirlik düzeylerinin belirlenmesi amacıyla yürütülmüştür. Farklı çalışmalarda her iki yaklaşım ölçek geliştirme süreci boyunca temel alınarak tüm sürecin iki yaklaşıma göre incelenmesi sağlanabilir. Ayrıca bu çalışma tipik performansın belirlendiği bir ölçme aracı için gerçekleştirilmiş, bilişsel ya da devinimsel Alana ilişkin ölçme araçlarında benzer çalışmalar yapıp, nitelikler incelenebilir.

KAYNAKÇA

- Baykul, Y. (2010). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması* (2. Baskı). Ankara: Pegem Akademi.
- Berberoğlu, G. (1989). Erişti testlerine madde seçiminde klasik test kuramı ve rasch modelinin karşılaştırılması. *Eğitim ve Bilim*, Cilt:13 S:74, 61-67.
- Can, S. (2003). *Ortaöğretim kurumları öğrenci seçme ve yerleştirme sınavı sözel bölümünün madde tepki kuramı modellerine göre analizi*. (Doktora tezi, Orta Doğu Teknik Üniversitesi, Ankara).
- Crocker, L. M., & Algina, J. (2006). Introduction to classical and modern test theory 2nd edth ed. *Victoria: Thomson Wadsworth*.
- Çelen, Ü. (2008). Klasik test kuramı ve madde tepki kuramı yöntemleriyle geliştirilen iki testin geçerlilik ve güvenilirliğinin karşılaştırılması, *İlköğretim Online*, 7(3), 758-768.
- Çelen Ü., ve Aybek, E. C. (2013). Öğrenci başarısının öğretmen yapımı bir testle klasik test kuramı ve madde tepki kuramı yöntemleriyle elde edilen puanlara göre karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 4(2), 64-75.
- Çelik, D. (2001). *Madde tepki kuramının (mtk) bir-, iki-, ve üç parametrelili modellerinin milli eğitim bakanlığı ortaöğretim kurumları öğrenci seçme ve yerleştirme sınavı testi verilerine uygunluğu*. (Doktora tezi, Orta Doğu Teknik Üniversitesi, Ankara).
- Christensen, K. B., & Kreiner, S. (2010). Monte Carlo tests of the Rasch model based on scalability coefficients. *British Journal of Mathematical and Statistical Psychology*, 63(1), 101-111.
- Courville, T. G. (2005). *An empirical comparison of item response theory and classical test theory item /person statistics*. (Doctoral dissertation, A&M University, Texas.)
- De Koning, E., Sijtsma, K., & Hamers, J. H. (2002). Comparison of four IRT models when analyzing two tests for inductive reasoning. *Applied Psychological Measurement*, 26(3), 302-320.
- DeMars, C. (2010). *Item response theory*. Oxford University Press.

- Egberink, I. J. A. L. (2010). *Applications of item response theory to non-cognitive data*. (Doctoral dissertation University of Groningen).
- Egberink, I. J., & Meijer, R. R. (2011). An item response theory analysis of Harter's Self-Perception Profile for Children or why strong clinical scales should be distrusted. *Assessment, 18*(2), 201-212.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. New Jersey: Lawrence Erlbaum Associates.
- Erkuş, A. (2003). *Psikometri üzerine yazılar*. Ankara: Türk Psikologlar Derneği Yayınları.
- Fan, X. (1998). Item response theory and classical theory: an empirical comparison of their item-person statistic. *Educational and Psychological Measurement, 58*(3), 357-381.
- Field, A. (2005). *Discovering statistics using SPSS*. Thousand Oaks, CA, US.
- Fraenkel, J. R., & Wallen, N. E. (2006). *How to design and evaluate research in education 2006*. Mc Graw Hill.
- Hambleton, R.K. (1990). Item response theory: introduction and bibliography. *Psicothema, 2*(1), 97-107.
- Hambleton, R.K. (1994). Item Response theory: a broad psychometric framework for measurement advances. *Psicothema, 6*(3), 535-556.
- Hambleton, R.K., Swaminathan H., and Rogers, H.J. (1991). *Fundamentals of item response theory*. California: Sage Publications Inc.
- Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Education Measurement: Issues and Practice, 12*(3), 38-47.
- Junker, B. W., & Sijtsma, K. (2001). Nonparametric item response theory in action: An overview of the special issue. *Applied Psychological Measurement, 25*(3), 211-220.
- Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement, 25*(3), 258-272
- Kılıç, İ. (1999). *Madde tepki kuramının (MTK) bir, iki ve üç parametrelili modellerinin Öğrenci Seçme ve Yerleştirme Merkezi'nin (ÖSYM) Öğrenci Seçme Sınavına (ÖSS) uygunluğu*. (Doktora tezi, Orta Doğu Teknik Üniversitesi, Ankara).
- Kuijpers R. E., van der Ark L.A., Croon, M.A.(2013). Standard errors and confidence intervals for scalability coefficients in Mokken scale analysis using marginal models. *Social Methodology, 43*:42–69.
- Ligtvoet R, van der Ark LA, Marvelde JM, Sijtsma, K. (2010). Investigating an Invariant Item Ordering for Polytomously Scored Items. *Educational Psychological Measurement, 70*:578–595.
- Linn, R.L. (1998). *Educational measurement*. New York: Macmillan Publishers.
- Loevinger, J. (1947). A systematic approach to the construction and evaluation of tests of ability. *Psychological Monographs, 61*. Washington, DC: American Psychological Association.
- Loevinger, J. (1948). The technique of homogeneous tests compared with some aspects of scale analysis and factor analysis. *Psychological Bulletin, 45*, 507–530.
- McDonald, R.P. & Paunonen, S. (2002). A Monte Carlo Comparison of Item and Person Statistics based on Item Response Theory versus Classical Test Theory. *Journal of Educational and Psychological Measurement, 62*, 921-943.
- Meijer, R. R. (2010). A comment on Watson, Deary, and Austin (2007) and Watson, Roberts, Gow, and Deary (2008): How to investigate whether personality items form a hierarchical scale?. *Personality and Individual Differences, 48*(4), 502-503.
- Meijer, R. R., Egberink, I. J., Emons, W. H., & Sijtsma, K. (2008). Detection and validation of unscalable item score patterns using item response theory: an illustration with Harter's Self-Perception Profile for Children. *Journal of Personality Assessment, 90*(3), 227-238.
- Molenaar, I. W. (2001). Thirty years of nonparametric item response theory. *Applied Psychological Measurement, 25*(3), 295-299.
- Özkurt, S. (2002). Madde Tepki Kuramı'nın (MTK) bir-, iki-, ve üç- parametrelili modellerinin bir İngilizce yeterlik başarı testi verilerine uygunluğu. (Doktora tezi, Orta Doğu Teknik Üniversitesi, Ankara).
- Progar, Š., Sočan, G., & Peč, M. (2008). An empirical comparison of item response theory and classical test theory. *Horizons of Psychology, 17*(3), 5-24.
- Reise, S.P., Ainsworth, A.T., and Haviland, M.G. (2005). Item Response theory. Fundamentals, applications, and promise in psychological research. *Current Directions in Psychological Science, 14*(2), 95-101.
- Sijtsma, K., & Molenaar, I. W. (2002). *Introduction to nonparametric item response theory* (Vol. 5). Sage.
- Sijtsma, K., Meijer, R. R., & van der Ark, L. A. (2011). Mokken scale analysis as time goes by: An update for scaling practitioners. *Personality and Individual Differences, 50*(1), 31-37.
- Sijtsma, K. & van der Ark, A. (2017). A tutorial on how to do a mokken scale analysis on your test and questionnaire data. *British Journal of Mathematical and Statistical Psychology, 70*, 137-158.
- Stage, C. (1998). A comparison between item analysis based on Item Response Theory and Classical Test theory. A study of the SweSAT Subtest WORD. *Educational Measurement, 29*.
- Van Schuur, W. H. (2011). *Ordinal item response theory: Mokken scale analysis* (Vol. 169). Sage.
- Yapar, T. (2003). *İki parametrelili tepki kuramı (MTK) modelinin yetenek kestirimleriyle Başkent Üniversitesi İngilizce yeterlik sınavının yordama geçerliğini inceleme çalışması*. (Doktora tezi, Orta Doğu Teknik Üniversitesi, Ankara.)

- Yalçın, M. (1999). *Eğitimi araştırma ve geliştirme dairesi başarı testlerinin madde-tepki kuramının bir, iki, üç parametreye uygunluğu*. (Doktora tezi, Orta Doğu Teknik Üniversitesi, Ankara).
- Yeğın, O. P. (2003). *Başkent Üniversitesi İngilizce yeterlik sınavının (Büiys) madde Madde Tepki Kuramı'nın (MTK) üç parametrelı modelinin kullanımıyla elde edilen yetenek kestirimlerinin yordama geçerliđi*. (Doktora tezi, Orta Dođu Teknik Üniversitesi, Ankara).