



Parametrik ve Parametrik Olmayan Madde Tepki Kuramı Modellerinden Çeşitli Faktörlere Göre Elde Edilen Madde ve Yetenek Kestirimlerinin Karşılaştırılması*

The Comparison of Item And Ability Estimations Calculated From The Parametric and Non-Parametric Item Response Theory According to The Several Factors

Ezgi Mor Dirlik, Kastamonu Üniversitesi, emor@kastamonu.edu.tr ORCID: 0000- 0003-0250-327X
Nizamettin Koç, Ankara Üniversitesi, nkoc@ankara.edu.tr ORCID: 0000-0002-3308-7849

Öz. Bu çalışmada, Madde Tepki Kuramı kapsamında yer alan iki farklı yaklaşım olan Parametrik Madde Tepki Kuramı (PMTK) ve Parametrik Olmayan Madde Tepki Kuramı (POMTK) modellerinden kestirilen madde ve yetenek parametrelerinin test uzunluğu, örneklem büyüklüğü ve maddelerin psikometrik nitelikleri faktörlerine göre karşılaştırılması amaçlanmıştır. Temel araştırma niteliğinde olan araştırmanın verileri Uluslararası Matematik ve Fen Eğilimleri Araştırması'nın (Trends of International Mathematics and Science-TIMSS) 2011 yılındaki uygulamasından elde edilmiştir. Çeşitli kriterler dikkate alınarak oluşturulan veri setinden çalışmanın amaçları ve incelemeye alınan faktörlerin koşulları gözetilerek 16 ayrı veri seti oluşturulmuştur. Elde edilen sonuçlar incelendiğinde, her iki yaklaşıma göre farklı örneklem büyüklüğünden kestirilen madde parametreleri arasında yüksek düzeyde ve manidar bir uyum olduğu belirlenmiştir. Test uzunluğu ve örneklem büyüklüğü koşullarının birlikte incelendiği veri setleri için ise PMTK ve POMTK modellerinden kestirilen madde parametreleri arasında yüksek düzeyde ve manidar ilişkiler olduğu bulunmuştur. Özetle her iki yaklaşıma göre farklı koşullar altında kestirilen madde ve yetenek parametreleri birbirleri ile yüksek düzeyde ve manidar ilişkiler göstermiştir. Böylelikle PMTK modellerinin varsayımları veri setleri için yeterli düzeyde karşılanmadığında, incelenen koşullara uygun durumlarda POMTK yaklaşımının tercih edilebileceği sonucuna ulaşılmıştır.

Anahtar Sözcükler: Parametrik madde tepki kuramı, parametrik olmayan madde tepki kuramı, Mokken ölçekleme, örneklem büyüklüğü, test uzunluğu.

Abstract. This study aimed to compare the item and ability parameters estimated from two different approaches of Item Response Theory, which are parametric and non-parametric, according to the factors of test length, sample size and the item psychometric qualities. The study was prepared in a basic research method and the data of the study is obtained from the Trends of International Mathematics and Science (TIMSS) 2011 application. From the main data set, 16 different data sets were formed according to the research questions and the conditions of the factors. The results showed that all of the item parameters estimated according to the parametric and non-parametric models from the data sets which are composed of different number of items and sample sizes were correlated significantly and highly. Among of the abilities estimated by non-parametric models in all sample size and test length factors, it was concluded that ability parameters are significantly and highly correlated. To sum up, it was found that both approaches had resulted in so highly and significantly correlated item and ability parameters. This finding showed that if the assumptions of parametric item response theory models are not met as a required level, the non-parametric alternative models can be used in the conditions that studied in this research.

Key words: Parametric item response theory, non-parametric item response theory, Mokken scaling, sample size, test length.

*Bu çalışma Ankara Üniversitesi Eğitim Bilimleri Fakültesi Ölçme ve Değerlendirme Anabilim Dalında 2017 ilk yazar tarafından sunulan aynı isimli doktora tezinin özeti niteliğindedir ve 5-8 Eylül 2018 tarihlerinde Kosova Prizren'de düzenlenen 6. Uluslararası Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresinde sözlü bildiri olarak sunulmuştur.

SUMMARY

Introduction

Measurement of the abstract constructs has been one of the most difficult fields since the beginning of the history of social sciences. There have been so many attempts to define and measure the properties such as success, motivation, intelligence, attitudes and personality. Because of the complex structure of these constructs, it has been already accepted that there is always a fair amount of error in these measurements. However, in order to minimize the level of error, the theories of measurement have been developed. The first measurement theory which has been known the most is the Classical Test Theory (CTT) and has been used so many measurement applications thanks to easy structural features'. Its assumptions can be met easily for very kind of data sets and this is the main reason that made this theory so popular. Despite of the popularity, CTT has some important drawbacks which limit the precision of the measurement. By using CTT, only one value of the error, the standard error of measurement, can be obtained but it is a fact that the amount of the error is not same for the all people in a sample. Also, it is so difficult to adapt a test to the tester level in this theory. In order to propose solutions to these drawbacks, the other measurement theory has been developed which is Item Response Theory (IRT). It is known as latent trait theory and provides deeper knowledge not only about the testees' performances but also the test items and the constructs. It has made possible to create adaptive tests and to equate tests for the goal of comparison of students' performance. However, in order to get these advantages of the theory, there are some requirements which should be met by the data sets. These requirements are large sample size, normal distribution, local independence and uni-dimensionality. Especially for the classroom assessment, it is not possible to reach large samples such as 1000 or 2000 generally and to apply as many items as in the high stake tests. For these reasons, the advantages of IRT cannot be used for the small samples which are not normally distributed. Because of the difficulty of these assumptions of IRT, it has been occurred new approaches which are based on this theory but differentiate as for requirements. One of this new approach is Nonparametric Item Response Theory, which is between CTT and Parametric Item Response Theory. It can be applied for short tests and small samples like CTT but it is possible to get invariable item and ability parameters like IRT. The nonparametric item response theory is a relatively new approach and has been frequently used in health sciences. Because of the novelty of this field, there are limited numbers of studies in literature and most of them have been focused on the theoretical perspectives. In this study, the comparison of the parameters calculated from the parametric and nonparametric IRT has been purposed.

Methods

By this study, it was aimed to compare the item and ability parameters estimated from two different approaches of Item Response Theory, which are parametric and non-parametric, according to the factors of test length, sample size and the item psychometric qualities. The study was prepared in a basic research method and the data of the study is obtained from the Trends of International Mathematics and Science (TIMSS) 2011 application. The mathematic booklet at eight grade level which had been determined as unidimensional at most was used and the data from the first 20 countries were gathered. The final data set of the research was composed of 7254 students from different countries. From the main data set, 16 different data sets were formed according to the research questions and the conditions of the factors. As for sample size factor, three different data sets were prepared consisted of 500, 1000 and 3000 students and sample size conditions were crossed with the test length conditions which are 5, 15 and 25 items. From these data sets, item and ability parameters were estimated by using the compatible parametric and non-parametric item response models and the results were compared by using the relevant correlation coefficients. In order to analyze item psychometric quality factor's effect on the parameters, four different data sets were composed and ability parameters were estimated from them and compared. Lastly, the reliability of the data sets was investigated by using test information functions in parametric item response theory models and

calculating Cronbach Alpha, Lambda2, Latent Class Reliability Coefficient(LCRC) and Molenaar and Sijtsma (MS) coefficients in non-parametric item response theory models.

Results

The results showed that all of the item parameters estimated according to the parametric and non-parametric models from the data sets which are composed of different number of items and sample sizes were correlated significantly and highly. However, as for ability parameters, it was determined that there is not consistency in the estimations according to the parametric models especially when the data sets are composed of five and 15 items. However, it was found that the abilities estimated from 25 item data set are correlated significantly and highly. Among of the abilities estimated by non-parametric models in all sample size and test length factors, it was concluded that ability parameters are significantly and highly correlated. Additionally, when the correlation of estimated ability parameters was examined between parametric and non-parametric models, parametric models produced consistent with non-parametric models only in the conditions of having 25 items or sample size is larger than 3000.

Discussion and Conclusion

In the scope of this work, the another factor whose effects were examined on ability parameters were the levels of item discrimination and item difficulty parameters and according to this factor, it was observed that the abilities estimated from the four data sets that were grouped in terms of 'low and high' item difficulties and item discrimination, were correlated highly and significantly. Accordingly, it was found that both approaches had resulted in so highly and significantly correlated item and ability parameters. This finding showed that if the assumptions of parametric item response theory models are not met as a required level, the non-parametric alternative models can be used in the conditions that studied in this research.

GİRİŞ

1930'lu yılların sonlarında öncüsü olan Klasik Test Kuramı (KTK)'nın sınırlılıklarını ortadan kaldırdığı iddiasıyla ortaya atılan Madde Tepki Kuramı (MTK) ile bireyin testle ölçülen gizil özellikleri ile test maddelerine verdiği tepkiler arasında bir örüntü kurulmakta ve bireyin Klasik Test Kuramında gerçek puan olarak tanımlanan yeteneği kestirilmektedir. Madde Tepki Kuramı sayesinde test maddesi ile testi alan bireyin özellikleri ilişkilendirilmekte ve söz konusu ilişki matematiksel olarak ifade edilmektedir (Hambleton ve Swaminathan, 1985). Bu kuram kapsamında test eşitleme, madde/test yanlılığını belirleme ve bireye uyarlanmış test geliştirme gibi karmaşık psikometri problemlerine etkili çözümler bulunmasına rağmen, bu çözümlerin anlamlılığı kuramın varsayımlarının karşılanma düzeyine bağlıdır (Embretson ve Reise, 2000). Varsayımların ihlal edilmesi durumunda MTK uygulamaları ciddi düzeyde yanlılığa ve yanlış kararlar vermeye neden olabilir. Kurama ilişkin varsayımların veri setleri için karşılanmaması durumunda uygulanabilecek iki yaklaşım vardır. Bu yaklaşımlardan biri tek boyutluluk varsayımının sağlanamadığı durumda uygulanmaktadır ve test ile ölçülen özellik çok boyutlu bir nitelik taşıyorsa, tek boyutlu madde tepki kuramı modellerine karşılık çok boyutlu madde tepki kuramı modelleri işe koşulmaktadır. Bir diğer alternatif yaklaşım ise, Parametrik Madde Tepki Kuramı (PMTK) modellerinin uygulanamadığı durumlarda Parametrik Olmayan Madde Tepki Kuramı (POMTK) modellerinin kullanımınıdır. Parametrik olmayan madde tepki kuramını, PMTK'dan ayıran temel özellik maddelerle ölçülen gizil yetenek arasındaki ilişkinin belirli bir modele bağlı kalınarak tanımlanmamasıdır. POMTK, madde tepki fonksiyonlarını herhangi bir parametrik fonksiyona bağlı kalmadan kestirmeye olanak tanır ve böylelikle, parametrik madde tepki kuramına göre daha fazla veri setine ve deneysel verilere daha iyi uyum sağlar (Sijtsma ve Junker, 2006).

Parametrik olmayan madde tepki kuramına dayalı olarak gerçekleştirilen ölçekleme tekniği alan yazında Mokken Ölçek Analizi (Mokken Scale Analysis) olarak da nitelendirilmekte ve test verilerini ölçeklemede sıkça kullanılmaktadır. Rasch ya da faktör analizi kadar popüler olmasa da, Google Akademik veri tabanında 5000'e yakın yayın kaydı bulunan bu teknik,

maddelerin ölçeklenmesinde PMTK modellerine göre daha esnek bir şekilde işlemekte ve parametrik modellerin doğrulayıcı yaklaşımından farklı olarak açımlayıcı bir yaklaşımı temel almaktadır (Straat, van der Ark ve Sijtsma, 2014). İkili puanlanan maddeler için Mokken (1971) tarafından geliştirilen bu yaklaşım, Molenaar (1991, 1997) tarafından çok kategorili puanlanan maddelere uyarlanmıştır. Ardından Sijtsma ve Molenaar (2001), POMTK'nın genel çerçevesini oluşturmuş, Meijer ve Baneke (2004) ise bu yaklaşımın PMTK'ya alternatif kullanımını örneklemiştir.

Ortaya atıldığı günden itibaren PMTK'ya göre daha az kısıtlayıcı olduğu ve verinin yapısı hakkında açımlayıcı bilgiler sağladığı için araştırmacıların ilgisini çeken POMTK, birçok farklı alanda kullanılmaktadır. Pazarlama (Pass ve Molenaar, 2003), sağlık ve tıp, siyasal bilimler (Van Schuur, 2003), psikoloji (Alterman, Cacciola, Habing ve Lynch, 2011) ve sosyoloji (Gow, Watson, Whitmean ve Dreary, 2011) bu ölçekleme türünün kullanıldığı alanlardandır (aktaran, Straat, van der Ark ve Sijtsma, 2014). Gelişen teknolojiyle paralel olarak kullanımı artan Mokken ölçekleme yaklaşımı Gutmann ölçeklemesi ile PMTK yaklaşımı arasında yer almaktadır. Mokken ölçekleme tekniği, testi alanların yanıt örüntüsünü tek ve gizil bir değişkenin göstergesi olarak kabul etmekte ve Gutmann ölçekleme yaklaşımının parametrik olmayan olasılıklı modelini sunmaktadır. Belirleyici (deterministic) bir yaklaşımdan ziyade olasılıksal bir yaklaşıma sahip olması ve gizil değişkenlerin çok az sayıda madde ile temsil edilmesinin zorunlu olduğu durumlarda dahi uygulanabilmesi bu yaklaşımın diğer ölçekleme tekniklerine göre avantajlarından (Van Schuur, 2011). Faktör analizi ve güvenilirlik analizleri maddeleri paralel olarak kabul edip ölçekleme yapmakta iken, Mokken ölçekleme tekniği, 1940'lı yıllarda Ferguson (1941) ve Carrol (1945) tarafından da belirtildiği gibi, maddelerin frekans dağılımının farklı olduğunu kabul etmekte ve bu sayede faktör analizinde zor olan ikili puanlanan verilerin analizlerini olanaklı hale getirmektedir (aktaran, Sijtsma ve Molenaar, 2002).

Her ne kadar parametrik olmayan bir yaklaşım olarak da nitelendirilse, bir veri setinin POMTK kapsamında ölçeklenebilmesi için bir takım varsayımların veri setleri tarafından karşılanması gerekmektedir. Bu varsayımlar tek boyutluluk, yerel bağımsızlık, madde tepki fonksiyonlarının monotonluğu ve kesişmeyen madde tepki fonksiyonlarıdır (Sijtsma ve Molenaar, 2002). Bu varsayımlardan tek boyutluluk ve yerel bağımsızlık parametrik ve parametrik olmayan madde tepki kuramı modellerinin ortak varsayımdır. POMTK ile PMTK arasındaki temel fark ise; POMTK'nın madde tepki fonksiyonlarına ilişkin PMTK kadar sınırlayıcı varsayımlarının olmamasıdır. Bu durum da verinin asıl yapısının korunmasını ve ölçülen psikolojik değişkenin daha iyi anlaşılmasını sağlamaktadır (Molenaar, 1997; Sijtsma ve Molenaar, 2002).

POMTK modellerinin madde tepki fonksiyonlarına ilişkin varsayımlarından ilki madde tepki fonksiyonlarının monotonluğudur. Bu varsayıma göre bireylerin maddeyi doğru yanıtlama olasılıkları ile ölçülen gizil yeteneğin arasındaki ilişkinin monoton ve azalmayan yönde olması gerekmektedir. Eğitimsel ve psikolojik testler gözetildiğinde, bu özellik şu şekilde yorumlanabilir: yeteneği, tutumu ya da becerisi daha üst düzeyde olan bireylerin söz konusu ölçülen özelliğe daha düşük düzeyde sahip olan bireylere göre maddelere doğru yanıt verme olasılıkları daha fazla olmasıdır (Meijer, Sijtsma ve Smidt 1990; Sijtsma ve Molenaar, 2002). Madde tepki fonksiyonlarına ilişkin POMTK kapsamında incelenen bir diğer varsayım ise, değişmez madde sıralaması, bir diğer ismiyle ise, kesişmeyen madde tepki fonksiyonları varsayımdır. Söz konusu varsayımın veri setleri için karşılanmasının ön koşulu, tüm maddelere ilişkin kestirilen madde tepki fonksiyonlarının Eşitlik 1'de verildiği gibi sıralanmasıdır (Sijtsma ve Meijer, 2007; Sijtsma ve Junker, 2006).

$$P_1(\theta) \leq P_2(\theta) \leq \dots \leq P_K(\theta) \quad (1)$$

Eşitlik 1 incelendiğinde, güçlük düzeyi en yüksek olan madde 1, güçlük düzeyi ikinci olarak en yüksek olan madde 2 tarafından takip edilir ve tüm maddeler için bu ilişkiler bu şekilde devam eder. Bu varsayımın karşılanması için maddeler arasındaki bu ilişkinin belli yetenek düzeyi aralıkları (θ) için sağlanması gerekir. Burada vurgulanması gereken nokta, varsayımla maddeyi doğru yanıtlama olasılıklarının değil madde sıralamasının aynı olması gerekliliğidir. Bu özellik, teorilerin test edilmesi, teste ilişkin başlama ve bitirme kurallarının belirlenmesi,

değişen madde fonksiyonlarını ortaya çıkarılması ve özellikle sağlık alanlarında kullanılan tanı ve teşhis amaçlı testlerle yapılan ölçmeler gibi birçok durum için önemlidir (Sijtsma ve Molenaar, 2002).

Tek boyutluluk, yerel bağımsızlık, madde tepki fonksiyonlarının monotonluğu varsayımlarını karşılayan veri setleri POMTK kapsamında yer alan esnek bir model olan Monoton Homojenlik Modeline (MHM) uyum sağlarken, bu varsayımlara ek olarak, değişmez madde sıralaması varsayımını karşılayan veri setleri ise daha kısıtlayıcı bir model olan Çift Monotonluk Modeline (ÇMM) uyum sağlar. Monoton homojenlik modelinde yer alan varsayımların veri seti için karşılandığı durumda, bireylerin toplam puanları gözetilerek yetenek düzeylerine göre stokastik (raslantısal) olarak sıralanabilir (aktaran, Straat, van der Ark ve Sijtsma, 2014). Monoton homojenlik modelinde yer alan varsayımlara ek olarak, kesişmeyen madde tepki fonksiyonları varsayımının da karşılanması gereken ÇMM 'de maddeler de yetenek düzeyleri gibi güçlüklerine göre sıralanabilir ve bu sayede MHM'den farklı olarak yalnızca maddeler değil, bireyler de aynı yetenek sürekliliğinde sıralanabilir. Bu özellik PMTK modelleri içinde yer alan Rasch modelde sağlanan bir özelliktir (Sijtsma ve Molenaar, 2002).

POMTK kapsamındaki modeller sadece ölçek düzeyinde değil madde düzeyinde de analizler yapmaya olanak sağlar. Her iki model için de geçerli olarak ölçme yapan maddelerin belirlenmesi için Mokken (1971, 1997), ilk olarak Loevinger'in (1947, 1948) kullandığı H ölçeklenebilirlik katsayısının kullanılmasını önermiştir. H katsayısı Guttman ölçeğindeki hata kavramıyla ilişkilidir ve bireyin zor maddeye doğru yanıt verirken kolay maddeye yanlış yanıt verme düzeyine ilişkin bilgi verir. H katsayısı bireyleri ölçülen özelliğe göre sıralamada maddenin ne kadar uyumlu olduğunu gösterir ve POMTK kapsamında "ölçeklenebilirlik katsayısı" olarak da adlandırılır (Sijtsma ve Molenaar, 2002). Ölçeklenebilirlik katsayısının üç türü vardır: maddeler için H katsayısı (H_i), madde çiftleri için H katsayısı (H_{ij}) ve ölçek için H katsayısı (H) olmak üzere üç tür ölçeklenebilirlik katsayısı hesaplanır. Teorik olarak H değerinin 0 ile 1 aralığında değer alması beklenir ve pozitif olan tüm H değerleri kabul edilebilir düzeydedir. H değerinin 1 olması ise Guttman hatasının 0'a yaklaştığını belirtir fakat uygulamada pek gerçekleşmeyen bir durumdur. Pratikte 0.0-0.3 aralığında H_i değerine sahip olan maddeler pozitif fakat düşük ayırt ediciliğe sahip maddeler olarak kabul edilir ve bireyleri toplam puana dayanarak güvenilir olarak sıralamaya katkısı düşük olduğundan test için kullanışsız olarak nitelendirilir (Mokken ve Lewis, 1982). Son zamanlarda yapılan bir çalışma da ise, Mokken tarafından önerilen bu ölçek analizinin yalnızca POMTK modelleri için değil, PMTK modelleri için de analize başlamadan önce yapılmasının faydalı olduğunu ve bu incelemenin model veri uyumunu arttıracakları öne sürülmüştür (Christensen ve Kreiner, 2010). Mokken tarafından önerilen bu ölçek analizinin özellikle küçük örneklerde etkili bir ölçek geliştirme stratejisi sunduğu da alan yazında yer almaktadır (Junker, 2000).

POMTK alanı gelişmeye devam eden bir alandır ve özellikle Hollanda'da ve Amerika Birleşik Devletleri'nde, POMTK modelleri, varsayımları ve uygulamaları sıkça araştırılan konulardan olmuştur. Gelişmeye devam eden bir alan olduğu için literatürde, uygulamalı çalışmalardan çok, kuramsal çalışmalar yer almaktadır. Kuram kapsamında güvenilirlik incelemelerini madde ve ölçek düzeyinde inceleyen çalışmalar alana ilişkin yapılan ilk çalışmalardandır (Molenaar ve Sijtsma, 1984; Meijer, Sijtsma ve Molenaar, 1995). Bu çalışmalarda KTK kapsamındaki madde ayırt edicilik parametresine karşılık ölçeklenebilirlik katsayılarının kullanılabilmesi belirtilmiş ve test düzeyinde güvenilirlik bilgisi veren MS katsayısı tanıtılmıştır. Farklı çalışmalarda PMTK ile POMTK'dan elde edilen madde parametreleri kıyaslanmış ve PMTK ile ölçeklenemeyen maddelerin POMTK kapsamında ölçeklenebildiği belirlenmiştir (Meijer, Egberink, Emons ve Sijtsma, 2008). PMTK ile farkların incelendiği bir çalışmada ise, POMTK kapsamında kullanılan Otomatik Madde Seçim İşlemi (OMSİ-Automated Items Selection Algorithm) ile farklı örneklem büyüklüklerinden elde edilen sınıflama oranları karşılaştırılmış ve madde düzeyindeki ölçeklenebilirlik katsayılarının 0.30 ve üstü seçilmesi durumunda örneklem büyüklüğünün 250 olduğu durumlarda dahi etkili sınıflamanın mümkün olduğu, ancak ölçeklenebilirlik katsayılarının düşük seçilmesi durumunda ise, gerekli örneklem büyüklüğünün 1500'e kadar çıkabileceği belirlenmiştir (Straat, van der Ark ve Sijtsma, 2014). Meijer ve Baneke (2004) tarafından yapılan uygulamalı bir çalışmada, PMTK ve

POMTK'dan kestirilen madde parametreleri karşılaştırılmış ve PMTK ile ölçeklenemeyen maddelerin POMTK kapsamında ölçeklenebildiği ve özellikle tipik performansla ilişkin yapılan ölçmelerde hem ölçek hem de ölçülen yapı hakkında POMTK ile daha fazla bilgi elde edildiği sonucuna ulaşılmıştır. Dyehouse (2009) ise PMTK ile daha önceden ölçeklenen bir veri setini POMTK ile ölçeklediği çalışmada, maddelerin çoğunun POMTK modellerinden Monoton Homojenlik Modeline(MHM) uyum sağladığını belirlemiş ve genel model uyum indekslerinin PMTK kapsamında daha yüksek olduğunu ortaya koymuştur. Zhou (2011) tarafından yapılan çalışmada ise, Pozitif ve Negatif Sendrom Ölçeği (Positive and Negative Syndrome Scale) POMTK kapsamında ölçeklenmiş ve PMTK'nın örneklem büyüklüğünün çok olduğu zamanlarda tutarlı sonuçlar verdiğini, POMTK'nın ise ölçülen psikolojik yapı hakkında derinlemesine bilgi sağladığını belirlemiştir.

Yapılan literatür taramasında; alana ilişkin yurtdışında kuramsal ağırlıklı olmak üzere pek çok çalışma yapıldığı ancak Türkiye'de yapılan çalışmaların sınırlı sayıda olduğu görülmektedir. PMTK'ya ilişkin pek çok sayıda araştırma yapılmış olmasına rağmen, POMTK uygulamalarına ilişkin yapılan araştırma sayısının çok az olduğu sonucuna ulaşılmıştır. Koğar (2014) çalışmada, basit ve iki boyutlu yapılarda, çeşitli örneklem büyüklükleri, test uzunlukları ve boyutlar arası korelasyon değerlerinde, tek boyutlu POMTK, tek boyutlu PMTK ve çok boyutlu MTK elde edilen madde parametreleri, maddelere ait model veri uyumları ve teste ait model veri uyumlarını incelemiş ve tüm simülasyon düzeneklerinde tek boyutlu POMTK modellerinin düşük hatalar ürettiği sonucuna ulaşılmıştır. Ayrıca tek boyutlu PMTK'da a parametresinin, örneklem büyüklüğündeki ve test uzunluğundaki artış ile birlikte azalma eğilimi gösterdiği fakat hiçbir simülasyon koşulundaki değişimin, b parametresini etkilemediği belirlenmiştir. Şengül Avşar ve Tavşancıl (2017) tarafından yapılan çalışmada çok kategorili puanlanan maddeler farklı test koşullarında POMTK ile ölçeklenmiş ve maddelerin psikometrik niteliklerindeki değişimler incelenmiştir. Çalışmanın sonuçlarına göre farklı test koşullarında POMTK yaklaşıma göre çözümlenmelerin yüksek güvenilirlik değerlerine sahip olan testler ürettiği sonucuna ulaşılmıştır.

İlgili çalışmalar incelendiğinde, POMTK modellerinin PMTK modellerine tamamen alternatif olması değil, her iki yaklaşımın da gereken durumlarda birlikte kullanılmasının uygun olduğu sonucuna ulaşıldığı görülmektedir. Bir diğer ifadeyle, yapılan çalışmaların tümünde PMTK yerine her durumda POMTK'nın kullanılmasını değil, her iki yaklaşımın da kendine özgü özelliklerinden faydalanılarak(üstünlüklerinden yararlanılarak) yapılan çalışmaların geliştirileceği, ölçülen özelliğin daha iyi anlaşılacağı ve model veri uyumsuzluğu sorununa bir ölçüde çözüm bulunabileceği savunulmaktadır. Ayrıca POMTK modelleri kapsamında veri yapısına dayalı olarak kestirilen madde tepki fonksiyonlarının ölçülen yapıya ilişkin daha fazla bilgi verdiği sonucuna ulaşılmış ve bu alanda özellikle deneysel çalışmaların yapılması gerekliliği alanda öncü araştırmacılar tarafından önerilmiştir (Christensen ve Kreiner, 2010; Junker, 2000; Van Schuur, 2011; Sijtsma ve Molenaar, 2002). Özellikle uygulamada ortaya çıkabilecek çeşitli durumların model kestirimleri üzerindeki etkilerinin incelenmesi, araştırmacılara kendi çalışma durumlarına göre model seçmeleri ve olası model veri uyumsuzluğu durumunda alternatif yollara başvurmaları konusunda yönlendirici olabilecektir. Tüm bu nedenlerle, bu çalışma kapsamında tek boyutlu ve iki kategorili puanlanan verilerde, çeşitli örneklem büyüklükleri, test uzunlukları ve maddelerin değişen psikometrik özelliklerinde, POMTK ile PMTK yaklaşımlarından elde edilen madde parametrelerini ve birey yeteneklerini incelemek bu çalışmanın problemini oluşturmaktadır.

Bu çalışmanın temel amacı, iki kategorili puanlanan verilerde PMTK modellerinden bir, iki ve üç parametrelili lojistik model ile POMTK kapsamında yer alan Mokken ölçekleme modellerinin, örneklem büyüklüğü, test uzunluğu ve maddelerin psikometrik özellikleri faktörlerine göre kestirilen madde ve birey yetenek parametrelerinin karşılaştırılmasıdır. Bu genel amaç kapsamında araştırmada yanıt aranan sorular şu şekildedir:

Araştırma kapsamında farklı örneklem büyüklüğü ve test uzunluğuna göre oluşturulan veri setlerinin, PMTK ve POMTK modellerine uyumları nasıldır?

İncelenen veri setinde yer alan maddelerin, PMTK modellerinden bir-iki ve üç parametrelili lojistik modelden en iyi uyumu sağlayan model ile POMTK modellerinden uyumlu olduğu

belirlenen model kullanılarak kestirilen madde ayırt edicilik ve güçlük parametreleri, örneklem büyüklüğüne ve test uzunluğuna göre nasıl farklılaşmaktadır?

İncelenen veri setinde yer alan bireylere ilişkin PMTK modellerinden bir-iki ve üç parametrelili lojistik modelden en iyi uyumu sağlayan model ile POMTK modellerinden uyumlu olduğu belirlenen model kullanılarak belirlenen yetenek kestirimleri, maddelerin psikometrik niteliklerine, test uzunluğuna ve örneklem büyüklüğüne göre nasıl farklılaşmaktadır?

Çalışma kapsamında incelenen faktörlere ilişkin oluşturulan veri setlerinin ölçeklendiği yaklaşıma göre güvenilirlik düzeyleri nasıldır?

YÖNTEM

Bu çalışma kapsamında çeşitli faktörlere göre düzenlenen verilerle parametrik ve parametrik olmayan madde tepki kuramı modelleri kullanılarak elde edilen madde ve yetenek parametrelerinin karşılaştırılması amaçlanmıştır. Araştırma farklı MTK yaklaşım ve modellerinin çeşitli faktörlere göre karşılaştırılmasını ve böylelikle söz konusu yaklaşımların özelliklerini ortaya çıkarmayı amaçladığı için temel araştırma niteliği taşımaktadır.

Çalışma Grubu

Çalışmanın evreni TIMSS 2011 uygulamasına katılan tüm ülkeler ve tüm katılımcılardır. Söz konusu evrenin genişliği göz önüne alındığında, örnekleme yöntemleri kullanılarak bir örneklem belirlenmesi gerekliliği ortaya çıkmaktadır. Araştırmanın genel amacına uygun olarak seçkisiz olmayan örnekleme belirleme yaklaşımından, uygun örnekleme yöntemi kullanılarak incelenen faktörler ve koşullarına göre birden çok örneklem belirlenmiştir.

TIMSS 2011 uygulaması kapsamında başarı testlerine göre ilk 20'de yer alan ülkeler çalışma grubuna dâhil edilmiştir. Başarı sıralaması yüksek olan ülkelerin çalışma grubuna alınma nedeni, yanıt ranjinin ve doğru yanıtlanma oranlarının yüksek olduğu durumlarda, MTK kapsamında model veri uyumunun daha yüksek düzeyde sağlanıyor olmasıdır (Baker ve Kim, 2004). Söz konusu ülkelerden incelemeye alınan kitapçığı yanıtlayan 7242 kişiden oluşturulan temel veri seti, çalışmanın amaçları doğrultusunda farklı veri setlerine ayrılmıştır.

Çalışma kapsamında madde ve yetenek parametrelerine etkisi incelenen ilk faktör örneklem büyüklüğü faktörüdür ve bu faktör altında örneklem büyüklüğünün 500, 1000 ve 3000 olduğu veri setleri oluşturulmuştur. Çalışmada madde ve yetenek parametrelerine etkisi incelenen bir diğer faktör ise, test uzunluğudur ve bu faktör altında test uzunluğunun 5, 15 ve 25 olduğu koşullar örneklem büyüklüğü faktöründe yer alan koşullar ile çaprazlanarak dokuz farklı veri seti oluşturulmuştur. Son olarak ise maddelerin psikometrik nitelikleri faktörü incelemeye alınmış ve madde güçlük parametrelerinin yüksek ve düşük ve madde ayırt edicilik parametrelerinin yüksek ve düşük olduğu dört ayrı veri seti oluşturulmuş ve tüm bu veri setlerinden birey yetenekleri ve madde parametreleri kestirilmiştir.

Verilerin Elde Edilmesi

Bu çalışma kapsamında incelenmeye alınan veriler TIMSS 2011 uygulaması kapsamında elde edilen verilerdir ve söz konusu uygulamaya ait uluslararası veri tabanı kullanılarak elde edilmiştir (<http://timss.bc.edu/timss2011/international-database.html>). TIMSS veri setinin çok geniş olması nedeniyle, çalışma kapsamında incelemeye alınan veri seti testte yer alan madde sayısının çok olması ve tek boyutluluk varsayımının yüksek düzeyde karşılanması kriterleri gözetilerek belirlenmiştir. Madde sayısının 8. Sınıf başarı testlerinde daha fazla olduğu belirlenmiş ve yapılan boyutluluk analizlerinde Matematik alt testlerinin Fen Bilimleri testine göre tek boyutluluk özelliğini daha yüksek düzeyde karşıladığı belirlenmiştir. Matematik alt testi kapsamında tek boyutluluk varsayımını en yüksek düzeyde karşılayan kitapçığın 13 nolu kitapçık olduğu belirlenmiş ve Matematik alt testi başarı sırasına göre ilk 20'de yer alan ülkelerden söz konusu kitapçığı yanıtlayan bireylerin yanıt örüntüleri kullanılarak çalışmanın temel veri seti oluşturulmuştur.

Çalışmada İncelemeye Alınan Faktörler

Çalışma kapsamında incelemeye alınan faktörler; veri setlerinde yer alan birey sayısı, test uzunluğu ve maddelerin psikometrik nitelikleridir. Uygulamada bu faktörlerin PMTK ve POMTK modelleri ile yapılan madde parametre kestirimlerini, birey yeteneklerini ve madde tepki fonksiyonlarını etkilediği belirtilmektedir (Molenaar, 1997; Van Schuur, 2011; Sijtsma ve Molenaar, 2002).

Araştırmada ele alınan ilk faktör gruplarda yer alan birey sayısıdır. Bu faktör için 500, 1000 ve 3000 kişilik veri setleri oluşturulmuştur. Bu örneklem büyüklüklerinin belirlenmesinde PMTK uygulamaları temel alınmıştır. 500 kişilik veri seti PMTK uygulamaları için küçük örneklem olarak nitelendirilirken, 1000 orta büyüklükte, 2000 ve 3000 ise büyük örneklem olarak nitelendirilmektedir. Büyük örneklem için 2000 yerine 3000 değerinin alınmasının nedeni, PMTK modellerinin 2000 kişilik bir örneklemde dahi, ölçülen psikolojik özelliğe ilişkin ranjının geniş olmadığı durumlarda, uç yetenek düzeylerinde yer alan bireylere ilişkin yetenek kestirimlerinin yüksek standart hata içermesinden dolayı alan yazında eleştiri almış olmasıdır. En küçük örneklem büyüklüğü için 500 kişilik veri setinin oluşturulma nedeni ise, POMTK modellerinin küçük örneklemelerde de etkili kestirimler yaptığı iddiasının test edilmesidir (Sijtsma ve Molenaar, 2002; Van Schuur, 2011). Alan yazında PMTK modelleri için yapılan analizlerde hatasız madde parametreleri kestirimi için en az 1000 kişilik veri setinin gerekli görüldüğü göz önünde bulundurularak (Mislevy ve Stocking, 1989), 500 kişilik veri setinin küçük ve 1000 kişilik veri setinin ise orta büyüklükte olarak nitelendirilebileceği düşünülmüştür.

Çalışma kapsamında incelenen bir diğer faktör ise test uzunluğudur. Bu faktör kapsamında, 5, 15 ve 25 maddelik testler incelemeye alınmış ve test uzunluğunun madde parametrelerine ve birey yetenek kestirimlerine etkisinin incelenmesi amaçlanmıştır. Bu faktörün çalışmaya dâhil edilme nedeni, POMTK modellerin testteki madde sayısının az olduğu durumlarda dahi birey yeteneklerine ve madde parametrelerine ilişkin PMTK modellerinden daha gerçekçi ve etkili kestirimlere olanak sağlamasıdır (Van Schuur, 2011; Sijtsma ve Molenaar, 2002). Söz konusu durumlarda PMTK ve POMTK kapsamında kestirilen parametreler arasındaki farkların incelenmesi gerçek uygulamalarda oluşabilecek koşullar ve elde edilecek kestirimler hakkında bilgi verici olacaktır. Çalışmada incelemeye alınan son faktör ise maddelerin psikometrik özellikleridir. Maddelerin güçlük ve ayırt edicilik özelliklerinin POMTK ve PMTK modelleri ile kestirilen yetenek parametrelerine etkisinin incelenmesi için madde güçlük ve ayırt edicilik değerlerinin yüksek ve düşük değerler aldığı maddeler belirlenmiş ve bu maddeler kullanılarak her iki yaklaşıma göre birey yetenekleri kestirilmiştir. Bu faktörün çalışmaya dâhil edilme nedeni ise özellikle ayırt edicilik ve güçlük değerleri normalden düşük ya da yüksek olan maddeler için PMTK modellerine göre POMTK modellerinin daha ayrıntılı kestirimlere olanak sağlamasıdır. Ayrıca bu faktör kapsamında söz konusu maddeler için POMTK'da daha gerçekçi MTF'lerin üretiliyor olması ve madde model uyumunun PMTK'ya göre daha yüksek düzeyde sağlanıyor olmasına ilişkin iddialar incelemeye alınmıştır (Junker, 2000). Söz konusu maddeler için her iki modelden kestirilen değerler arasındaki farkların incelenmesi yaklaşımların doğası açısından bilgi verici olacaktır. Belirtilen faktörlere göre toplam 16 farklı veri seti oluşturulmuştur ve analizler gerçekleştirilmiştir.

Verilerin Analizi

Çalışmanın temel amacına paralel olarak oluşturulan veri setleri parametrik ve parametrik olmayan madde tepki kuramı kapsamında analiz edilmiştir. Her iki yaklaşım için de öncelikle varsayımlar test edilmiştir. Tek boyutluluk varsayımı için PMTK kapsamında tetrakorik korelasyon matrisine dayalı olarak Açıklayıcı Faktör Analizi yapılırken, POMTK kapsamında ise Otomatik Madde Seçim Süreci (OMSİ) uygulanmıştır. Yerel bağımsızlık analizleri için her iki yaklaşımda da artık korelasyon matrisi oluşturulmuş ve bu matriste yer alan değerler incelenmiştir. Model veri uyumu incelemeleri ve parametre kestirimleri için ise PMTK kapsamında BILOG-MG programı ve R programında yer alan "irtos" ve "ltm" paketleri

kullanılırken, POMTK analizleri için yine R programında yer alan “mokken” ve “KernSmoothIrt” paketleri kullanılmıştır.

Çalışma kapsamında PMTK ve POMTK kestirimlerinde kullanılan teste ilişkin güvenilirliğin belirlenmesi için ise PMTK kapsamında test bilgi fonksiyonları oluşturulmuş ve marjinal güvenilirlik katsayıları hesaplanmıştır. POMTK ile incelenen veri setleri için yapılan güvenilirlik analizinde ise, Lambda 2 ve Cronbach Alfa katsayılarına ek olarak alana özgü LRCR ve MS katsayıları da hesaplanmıştır.

POMTK'ya göre yapılan ölçeklemelerde alan yazında belirtilen izlenmesi gereken adımlar takip edilmiş ve veri setlerinin Çift Monotonluk Modeline uyumu incelenmiştir. Bu duruma neden olarak da Çift Monotonluk Modelinin Monoton Homojenlik Modelini kapsadığı ve Çift Monotonluk Modeline uyumlu olan veri setinin Monoton Homojenlik Modeline zaten uyumlu olmak zorunda olduğunu belirtmektedirler(Junker, 2000; Meijer, Tenderio ve Wanders, 2015).

BULGULAR

Araştırmadan elde edilen bulgular yanıt aranan araştırma sorularının sıralaması gözetilerek bu bölümde verilmiştir.

İlgili Faktörlere Göre Oluşturulan Veri Setlerinin PMTK ve POMTK'ya Uyularının İncelenmesi

Çalışma kapsamında yanıt aranan ilk soru, farklı faktörlere göre oluşturulan veri setleri için uyumlu PMTK ve POMTK modellerinin belirlenmesidir. Bu amaçla ilk olarak örneklem büyüklüğü, test uzunluğu ve maddelerin psikometrik nitelikleri faktörlerinde yer alan koşullara göre oluşturulan veri setlerinin PMTK ve POMTK'ya olan uyumları incelenmiştir. PMTK'ya ilişkin yapılan model veri uyumu analizleri sırasında -2Loglikelihood değerleri ile maddelerin ki-kare istatistikleri göz önünde bulundurulurken, POMTK'ya ilişkin yapılan model veri uyumu incelemelerinde ise, alan yazında önerilen analiz adımlarına bağlı kalınarak, ilk olarak madde güçlükleri ve ölçeklenebilirlik katsayıları hesaplanmıştır. Bunun ardından madde tepki fonksiyonlarının monotonluğu ve kesişmezliği varsayımları incelenerek model veri uyumu incelemeleri yapılmış ve veri setleri için uygun POMTK modeli belirlenmiştir. Elde edilen sonuçlar çalışmada incelemeye alınan faktörler ve koşullarına göre Tablo 1'de verilmektedir.

Tablo 1'de yer alan veri setleri incelendiğinde, örneklem büyüklüğü faktörü altında 30 maddeden oluşan 500, 1000 ve 3000 kişilik veri setinin incelemeye alındığı görülmektedir. Test uzunluğu ve örneklem büyüklüğü faktör koşullarının çaprazlanması ile elde edilen dokuz ayrı veri setinin incelemeye alındığı ve son olarak da maddelerin farklı psikometrik nitelikleri faktörü altında a ve b parametresinin yüksek ve düşük olarak ayrı ayrı incelemeye alındığı dört veri seti olduğu görülmektedir. Oluşturulan veri setlerine göre yapılan model veri uyumu incelemelerine bakıldığında, PMTK için çoğunlukla 2PLM'ye uygun olduğu belirlenmiştir. PMTK kapsamında yapılan model veri uyumu incelemelerinde alan yazında belirtildiği üzere örneklemde yer alan birey sayısındaki artışın model veri uyumunu iyileştirdiği gözlenmiş ancak madde sayısının model veri uyumu üzerindeki etkisinin örneklem büyüklüğüne göre daha önemli olduğu belirlenmiştir. Beş maddeden oluşan 500 kişilik veri seti için hiçbir PMTK modeli uygun bulunmamış ancak analizlerde tutarlılığı sağlamak için kestirimler 2 PLM ile yapılmıştır. Maddelerin farklı psikometrik nitelikleri faktörü altında incelenen koşullara göre oluşturulan veri setlerinin PMTK modellerine uyumu incelendiğinde ise, a-yüksek olarak belirtilen madde ayırt edicilik parametresi yüksek maddelerden oluşan veri setinin 1PLM'ye, b-yüksek olarak nitelendirilen madde güçlük parametrelerinin yüksek olduğu veri setinin ise beklenildiği üzere şans parametresinin de kestirildiği 3PLM'ye uyum sağladığı belirlenmiştir.

Tablo 1. Veri setlerinin PMTK ve POMTK kapsamında model veri uyumu incelemeleri

Faktör	Koşullar	Uyumlu Model	
		PMTK	POMTK
Örneklem	500 kişi	2 PLM	MHM
Büyükülüğü	1000 kişi	2PLM	MHM
(Madde Sayısı 30)	3000 kişi	3PLM	MHM
Test	5 madde 500 kişi	Belirlenememiş	MHM
Uzunluğu*Örneklem	5 madde 1000 kişi	2PLM	MHM
Büyükülüğü	5 madde 3000 kişi	2PLM	MHM
	15 madde 500 kişi	2PLM	MHM
	15 madde 1000 kişi	2PLM	MHM
	15 madde 3000 kişi	2PLM	MHM
	25 madde 500 kişi	2PLM	ÇMM
	25 madde 1000 kişi	2PLM	ÇMM
	25 madde 3000 kişi	3PLM	ÇMM
Maddelerin Farklı	a-yüksek	1PLM	ÇMM
Psikometrik	a-düşük	2PLM	MHM
Nitelikleri	b-yüksek	3PLM	MHM
	b-düşük	1PLM	MHM

Farklı Örneklem Büyükülüğüne ve Test Uzunluğuna Sahip Veri Setlerinden PMTK ve POMTK Kapsamında Kestirilen Madde Parametrelerinin Karşılaştırılması

Örneklem büyükülüğü

Bir önceki araştırma sorusu kapsamında PMTK ve POMTK modellerine uyumları belirlenen veri setlerinden uyumlu oldukları model gözetilerek madde parametreleri kestirilmiş ve kestirilen parametreler arasındaki korelasyon katsayıları hesaplanmıştır. Madde düzeyinde yapılan incelemelerde öncelikli olarak madde ayırt edicilik parametreleri arasındaki korelasyonlar incelenmiştir. Parametreler için normallik varsayımı sağlanmadığı için Spearman Brown Korelasyon Katsayıları hesaplanmış ve Tablo 2’de verilmiştir.

Tablo 2. PMTK ve POMTK’ya göre kestirilen madde ayırt edicilik parametreleri arasındaki korelasyon katsayıları

Spearman Brown KK	PMTK			POMTK	
	500a	1000a	3000a	500H	1000H
1000a	0.880**				
3000a	0.938**	0.946**			
500H	0.958**	0.845**	0.890**		
1000H	0.821**	0.945**	0.899**	0.837**	
3000H	0.867**	0.904**	0.945**	0.894**	0.942**

**p<0.01

Tablo 2’de görülen veri setlerine ilişkin isimlendirmeler incelendiğinde, 500, 1000 ve 3000 örneklem büyükülüğünü belirtirken, a, PMTK kapsamında kestirilen madde ayırt edicilik parametrelerini, H ise POMTK’ya göre kestirilen madde ayırt edicilik parametrelerini göstermektedir. Buna göre hesaplanan korelasyon katsayıları incelendiğinde, tüm örneklemere göre kestirilen madde parametreleri arasındaki ilişkilerin yüksek ve manidar olduğu görülmektedir. En yüksek ilişki 500 kişilik veri setlerinden kestirilen parametreler arasında 0.958 olarak bulunmuş, en düşük ilişki ise 500 ve 1000 kişilik veri setleri arasında 0.821 olarak hesaplanmıştır. Ayrıca tüm örneklem büyükükleri için en yüksek ilişkiler aynı büyükükteki örneklemeler arasında hesaplanmıştır.

Madde ayırt edicilik parametrelerine ek olarak hesaplanan diğer madde parametresi madde güçlük indeksidir ve farklı örneklem büyüklüklerinden kestirilen madde güçlük parametreleri arasındaki ilişkiler Tablo 3'te verilmiştir.

Tablo 3. PMTK ve POMTK'ya göre kestirilen madde güçlük parametreleri arasındaki korelasyon katsayıları

Spearman Brown KK	PMTK			POMTK	
	500b	1000b	3000b	500	1000
1000b	0.994**				
3000b	0.994**	0.996**			
500	-0.979**	-0.974**	-0.968**		
1000	-0.977**	-0.981**	-0.974**	0.993**	
3000	-0.977**	-0.977**	-0.976**	0.994**	0.996**

**p<0.01

Tablo 3'te 500b, 1000b ve 3000b olarak belirtilen değerler, 500, 1000 ve 3000 kişilik veri setlerinden PMTK'ya göre kestirilen madde güçlük indeksleri iken, 500, 1000 ve 3000 olarak belirtilen değerler ise POMTK yaklaşımı kapsamında Monoton Homojen Modelinden elde edilen parametreleri göstermektedir. Elde edilen sonuçlar incelendiğinde, her iki yaklaşımdan elde edilen parametreler arasındaki korelasyon katsayılarının, yüksek ve manidar olduğu görülmektedir. PMTK ve POMTK'ya göre kestirilen madde güçlük parametreleri arasındaki ilişkinin yönünün negatif olması ise, POMTK kapsamında madde güçlüğüne KTK'da olduğu gibi p (percent of correct) değeri olarak tanımlanmasından kaynaklanmaktadır.

Test uzunluğu

Bu faktör altında incelenen üç koşul olan, 5, 15 ve 25 maddelik veri setlerinden PMTK ve POMTK kapsamında uygun olduğu belirlenen modeller kullanılarak madde parametreleri kestirilmiş, kestirilen parametreler arasındaki ilişkiler hesaplanmış ve elde edilen sonuçlar test uzunluğu koşullarına göre sıralı olarak sunulmuştur.

5, 15 ve 25 maddelik veri setlerinden kestirilen madde ayırt edicilik ve güçlük parametreleri arasındaki ilişkiler Tablo 4, 5 ve 6'da verilmiştir.

Tablo 4. 5 maddelik veri setleri için PMTK ve POMTK modelleri ile hesaplanan madde ayırt edicilik parametreleri arasındaki korelasyonlar

Spearman Brown KK	PMTK			POMTK	
	500a	1000a	3000a	500H	1000H
1000a	1.00**				
3000a	1.00**	1.00**			
500H	0.900*	0.900**	0.900**		
1000H	0.800**	0.800**	0.800**	0.900**	
3000H	0.900**	0.900**	0.900**	1.00**	0.900**

**p<0.01

Tablo 4'te yer alan veri setlerinin isimlendirilmesinde öncelikli olarak veri setinde yer alan birey sayısı belirtilmiş, ardından PMTK modelleri ile kestirilen madde ayırt edicilik indeksleri *a* ile, POMTK modelleri ile hesaplanan madde ayırt edicilik indeksleri ise *H* ile gösterilmiştir. PMTK ve POMTK'dan kestirilen madde ayırt edicilik parametreleri arasındaki korelasyon katsayıları incelendiğinde, hem yaklaşımların kendi içinde hem de farklı yaklaşımlardan kestirilen parametreler arasındaki korelasyonların yüksek ve manidar olduğu görülmektedir. PMTK'dan kestirilen parametreler arasındaki ilişkinin ise mükemmel düzeyde, 1.00 korelasyon katsayısına sahip olduğu görülmektedir. Aynı durum POMTK kapsamında 500 ve 3000 kişilik veri setlerinden kestirilen parametreler için geçerlidir.

Tablo 5. 15 maddelik veri setleri için PMTK ve POMTK modelleri ile hesaplanan madde ayırt edicilik parametreleri arasındaki korelasyonlar

Spearman Brown KK	PMTK			POMTK	
	500a	1000a	3000a	500H	1000H
1000a	0.950**				
3000a	0.968**	0.975**			
500H	0.929*	0.883**	0.893**		
1000H	0.840**	0.885**	0.861**	0.921**	
3000H	0.893**	0.884**	0.909**	0.966**	0.957**

**p<0.01

Tablo 5'te yer alan veri setlerinin isimlendirilmesinde bir önceki tablodaki sistem gözetilmiş ve önce örnekleme yer alan birey sayısı ardından PMTK'dan kestirilen parametreler *a* ile POMTK'dan kestirilen parametreler ise *H* ile gösterilmiştir. 15 maddelik veri setlerinden kestirilen madde ayırt edicilik parametreleri arasındaki ilişkiler incelendiğinde, gerek yaklaşımlar içinde gerekse her iki yaklaşım kapsamında kestirilen parametreler arasındaki korelasyon katsayılarının oldukça yüksek ve manidar olduğu görülmektedir.

Tablo 6. 25 maddelik veri setleri için PMTK ve POMTK modelleri ile hesaplanan madde ayırt edicilik parametreleri arasındaki korelasyonlar

Spearman Brown KK	PMTK		POMTK	
	500a	1000a	3000a	500H
1000a	0.905**	1.00		
3000a	0.937**	0.876**	1.00	
500H	0.863**	0.966**	0.867**	1.00
1000H	0.882**	0.937**	0.915**	0.957**

**p<0.01

Tablo 6'da yer alan veri setleri de Tablo 4 ve 5'te yer alan veri setlerinde olduğu gibi isimlendirilmiştir. PMTK kapsamında 500 kişilik veri setine ilişkin katsayıların tabloda yer almama nedeni, bu veri setinin 1 PLM ile ölçeklenmesi dolayısıyla bu veri seti için madde ayırt edicilik parametresinin kestirilmemiş olmasıdır. Elde edilen korelasyon katsayıları incelendiğinde hem aynı yaklaşımdan hem de farklı yaklaşımlardan kestirilen madde ayırt edicilik parametreleri arasındaki ilişkinin oldukça yüksek ve manidar olduğu bulunmuştur. Hesaplanan korelasyon katsayıları test uzunluklarına göre incelendiğinde, 25 maddelik veri setlerinden kestirilen korelasyon katsayılarının 5 maddelik veri setlerinden kestirilen korelasyon katsayılarına göre yüksek ancak 15 maddelik veri setlerinden kestirilen korelasyon katsayılarına göre daha düşük olduğu bulunmuştur. Dolayısıyla PMTK ve POMTK kapsamında kestirilen madde ayırt edicilik indekslerinin en yüksek düzeyde uyumu 15 maddelik veri setlerinde gösterdiği sonucuna ulaşılabilir. Ayrıca en yüksek ilişkiler aynı örneklem büyüklüklerine sahip veri setlerinden kestirilen madde ayırt edicilik parametreleri için hesaplanmıştır.

Her iki yaklaşım kapsamında kestirilen diğer madde parametresi ise, madde güçlük parametresidir. Test uzunluğu ve örneklem büyüklüğü faktörleri koşulları kapsamında kestirilen madde parametreleri arasındaki korelasyonlar Tablo 7,8 ve 9'de verilmiştir.

Tablo 7. 5 maddelik veri setleri için PMTK ve POMTK modelleri ile hesaplanan madde güçlük parametreleri arasındaki korelasyonlar

Spearman Brown KK	PMTK			POMTK	
	500b	1000b	3000b	500	1000
1000b	1.00**				
3000b	1.00**	1.00**			
500	-0.975**	-0.975**	-0.975**		
1000	-1.00**	-1.00**	-1.00**	0.975**	
3000	-1.00**	-1.00**	-1.00**	0.975**	1.00**

**p<0.01

Tablo 7’de yer alan veri setlerinin isimlendirilmesinde öncelikle örnekleme yer alan birey sayısı belirtilmiş ve PMTK kapsamında kestirilen madde güçlük parametreleri *b* ile gösterilirken, POMTK için yalnızca örneklem büyüklüğü bilgisi kullanılmıştır. Kestirilen korelasyon katsayıları incelendiğinde hem yaklaşımlar içinde hem de yaklaşımlar arasındaki ilişki düzeylerin mükemmel yakın ve manidar olduğu görülmektedir. PMTK ile POMTK’den kestirilen madde güçlük parametreleri arasındaki ilişkinin yönünün negatif olması ise POMTK kapsamında madde güçlüğü KTK’da olduğu gibi *p* değeri olarak alınmasından kaynaklanmaktadır. Özetle 5 maddelik veri setlerinden kestirilen madde güçlük indeksleri arasındaki ilişki oldukça yüksek bulunmuştur.

Tablo 8. 15 maddelik veri setleri için hesaplanan madde güçlük parametreleri arasındaki korelasyonlar

Spearman Brown KK	PMTK			POMTK	
	500b	1000b	3000b	500	1000
1000b	0.982**				
3000b	0.964**	0.971**			
500	-0.928*	-0.950**	-0.928**		
1000	-0.987**	-0.965**	-0.945**	0.976**	
3000	-0.974**	-0.942**	-0.929**	0.983**	0.990**

**p<0.01

Tablo 8’de 15 maddelik veri setlerinden PMTK ve POMTK’ya göre kestirilen madde güçlük parametreleri arasında hesaplanan korelasyon katsayıları verilmiştir. Katsayılar incelendiğinde, hem yaklaşımların kendi içinde hem de karşılıklı olarak kestirilen parametreler arasındaki ilişkilerin mükemmel düzeyde ve manidar olduğu görülmektedir. Dolayısıyla 5 maddelik veri setlerinde olduğu gibi, test uzunluğunun 15 madde olduğu koşulda da PMTK ve POMTK kapsamında hesaplanan madde güçlük indeksleri arasında mükemmel ve manidar ilişkiler olduğu sonucuna ulaşılmıştır.

Tablo 9. 25 maddelik veri setleri için hesaplanan madde güçlük parametreleri arasındaki korelasyonlar

Spearman Brown KK	PMTK			POMTK		
	500b	1000b	3000b	500	1000	3000
500b	1.00					
1000b	0.991**	1.00				
3000b	0.982**	0.972**	1.00			
500	-0.987**	-0.975**	-0.987**	1.00		
1000	-0.983**	-0.980**	-0.991**	0.989**	1.00	
3000	-0.981**	-0.972**	-0.999**	0.989**	0.992**	1.00

**p<0.01

Tablo 9’da incelenen son test uzunluğu koşulu olan 25 maddelik veri setleri için PMTK ve POMTK’ya göre kestirilen madde güçlük parametreleri arasındaki ilişkiler verilmiştir. Elde edilen korelasyon katsayıları incelendiğinde, tüm koşullardan elde edilen madde güçlük parametreleri arasındaki korelasyonların oldukça yüksek ve manidar olduğu belirlenmiştir. Bu

bulgularından yola çıkarak, incelemeye alınan farklı test uzunluğu ve örneklem büyüklüğü koşulları için PMTK ve POMTK kapsamında kestirilen madde parametreleri arasında tüm koşullarda çok yüksek ve manidar ilişkiler olduğu sonucuna ulaşılmıştır.

Yetenek Parametrelerinin Farklı Örneklem Büyüklüğü ve Test Uzunluğu Koşulları ile Maddelerin Farklı Psikometrik Niteliklerine Göre Değişiminin İncelenmesi

Çalışma kapsamında ikinci araştırma sorusunda farklı test uzunluğu ve örneklem büyüklüğü koşullarına göre madde parametre kestirimleri yapılmıştır. Madde parametrelerinin belirlenmesinin ardından, yetenek parametrelerine ilişkin kestirimlere geçilmiş ve yetenek parametrelerinin maddelerin farklı psikometrik niteliklerinden, test uzunluğu ve örneklem büyüklüğü koşullarından nasıl etkilendiği incelenmiştir.

İncelemeye alınan ilk faktör maddelerin farklı psikometrik nitelikleridir. Model veri uyumunun PMTK ve POMTK'da en iyi düzeyde sağlandığı 1000 kişilik veri seti kullanılarak, madde güçlük ve ayırt edicilik parametreleri yüksek ve göreceli olarak düşük maddeler belirlenerek dört ayrı veri seti oluşturulmuştur. Tüm maddeler, her iki yaklaşıma göre kestirilen madde parametre değerleri gözetilerek, madde güçlük parametrelerine göre kolaydan zora ve ayırt edicilik parametrelerine göre ise düşükten yükseğe doğru sıralanmış ve bu veri setlerinin uyumlu oldukları PMTK ve POMTK modelleri belirlenerek yetenek kestirimleri yapılmıştır.

Madde ayırt edicilik parametresine göre yüksek ve düşük olan maddelerden oluşturulan veri setlerine ilişkin yapılan model veri uyumu incelemesinde, madde ayırt edicilik değerlerinin düşük olduğu veri setinin PMTK'da iki parametrelili lojistik modele, POMTK'da ise, monoton homojenlik modeline uyum sağladığı belirlenirken, ayırt edicilik değerlerinin yüksek olduğu veri setinin PMTK'da bir parametrelili lojistik modele, POMTK'da ise çift monotonluk modeline uyum sağladığı belirlenmiştir. Yetenek kestirimleri söz konusu modeller kullanılarak yapılmıştır ve kestirilen yetenek parametreleri arasındaki ilişkiler Kendall Tau Korelasyon Katsayısı kullanılarak incelenmiştir. Bu katsayının tercih edilme nedeni, kestirilen yetenekler için normal dağılım varsayımının sağlanmaması ve düşük ayırt edicilik değerlerine sahip maddelerden oluşan veri setinin POMTK kapsamında MHM ile ölçeklenmesinden dolayı yetenek parametrelerinin sıralama ölçeği düzeyinde elde edilmiş olmasıdır. Hesaplanan korelasyon katsayıları Tablo 10'da verilmiştir.

Tablo 10. *En yüksek ve en düşük madde ayırt edicilik parametrelerine sahip maddelerden oluşan veri setlerinden PMTK ve POMTK'ya göre kestirilen yetenekler arasındaki ilişkiler*

Kendall Tau KK	PMTK		POMTK
	En düşük	En yüksek	En düşük
PMTKenyüksek	0.849**		
POMTKendusuk	0.991**	0.840**	
POMTKenyüksek	0.849**	1.000**	0.840**

**p<0.01

Tablo 10'da verilen değerler incelendiğinde, düşük madde ayırt edicilik parametrelerine sahip maddelerden oluşan veri setinden her iki yaklaşıma göre kestirilen yetenekler arasında mükemmel ve manidar ilişkiler olduğu görülmektedir. Aynı durum en yüksek madde ayırt edicilik parametrelerine sahip veri setlerinden kestirilen yetenekler için de geçerlidir. Özetle, her iki yaklaşımdan aynı veri seti kullanılarak kestirilen yetenekler arasındaki ilişkinin, aynı yaklaşımla farklı veri seti kullanılarak yapılan yetenek kestirimleri arasındaki ilişkiden daha yüksek olduğu bulunmuştur. Bir diğer ifadeyle PMTK ile en düşük ayırt edicilik değerine sahip veri setinden kestirilen yetenek parametreleri ile POMTK ile aynı veri setinden kestirilen yetenekler arasındaki ilişki, PMTK ile en yüksek ayırt ediciliğe sahip maddelerden oluşan veri setinden kestirilen yetenek parametreleri arasındaki ilişkiden daha yüksektir.

İncelemeye alınan diğer madde niteliği ise madde güçlük parametresidir. Her iki yaklaşıma göre öncelikli olarak model veri uyumu incelemeleri yapılmıştır. Madde güçlük parametreleri yüksek olan maddelerden oluşturulan veri setinin PMTK modellerinden en iyi uyumu yedi madde ile üç parametrelili lojistik model için sağladığı belirlenmiştir. Madde güçlük düzeyi düşük olan maddelerden, bir diğer ifadeyle kolay maddelerden oluşan veri seti ise en iyi uyumu sekiz

uyumlu madde ile bir parametrelili lojistik model için sağlamıştır. Dolayısıyla en kolay ve zor maddelerden oluşan veri setleri için PMTK kapsamında bir ve üç parametrelili lojistik model kullanılarak yetenek kestirimleri yapılmıştır. PMTK ve POMTK'ya uyumlu olan modeller belirlendikten sonra, her iki veri seti için de iki yaklaşıma dayalı olarak yetenek kestirimleri yapılmıştır. POMTK ile yapılan kestirimler ile PMTK kullanılarak yapılan yetenek kestirimleri arasındaki ilişki Kendall Tau Korelasyon Katsayısı kullanılarak hesaplanmış ve elde edilen sonuçlar Tablo 11'de verilmiştir.

Tablo 11. *En yüksek ve en düşük madde güçlük parametrelerine sahip maddelerden oluşan veri setlerinden POMTK ve PMTK'ya göre kestirilen yetenekler arasındaki ilişkiler*

KendallTau KK	POMTK		PMTK
	En kolay	En zor	En zor
POMTKenzor	0.837**		
PMTKenzor	0.829**	0.975**	
PMTKenkolay	0.100**	0.837**	0.829**

**p<0.01

Tablo 11'de yer alan değerler incelendiğinde, her iki yaklaşımda da en zor ve kolay maddelerden kestirilen yetenekler arasındaki ilişki düzeylerinin benzer olduğu görülmektedir. Bu değerler POMTK için 0.837, PMTK için 0.829 olarak bulunmuştur. Dolayısıyla PMTK ve POMTK için, zor ve kolay maddelerden kestirilen yetenekler arasında manidar ve yüksek ilişkiler olduğu sonucuna ulaşılmaktadır. Yetenekler arasındaki ilişkiler PMTK ve POMTK yaklaşımlarına göre incelendiğinde ise, POMTK kapsamında en kolay maddelerden oluşan veri setinden kestirilen yetenekler ile aynı veri setinden PMTK ile kestirilen yetenekler arasında mükemmel ve manidar ilişkiler olduğu görülmektedir. Benzer durum en zor maddelerden oluşan veri setlerinden kestirilen yetenek parametreleri için de geçerlidir. POMTK ve PMTK'dan en zor maddeler kullanılarak yapılan yetenek kestirimleri arasında mükemmel yakın 0.975 ve manidar bir ilişki olduğu belirlenmiştir. Her iki yaklaşıma göre farklı veri setlerinden kestirilen yetenek parametrelerinin ilişkileri incelendiğinde ise, yine yüksek ve manidar ilişkiler (0.829 ve 0.837) olduğu görülmektedir. Özetle, veri setinde yer alan maddelerin güçlük düzeyi fark etmeksizin, her iki yaklaşımdan kestirilen yetenek parametreleri arasında yüksek ve manidar ilişkiler olduğu ve aynı veri setinden her iki yaklaşımla kestirilen yetenekler arasındaki ilişkilerin mükemmel yakın ve manidar olduğu belirlenmiştir.

Maddelerin farklı psikometrik niteliklerine göre oluşturulan veri setlerinden kestirilen yetenekler arasındaki ilişkilerin incelenmesinin ardından, farklı madde sayısı ve örneklem büyüklüklerinden oluşan veri setlerinden PMTK ve POMTK'ya dayalı yetenek kestirimlerinin yapılmasına geçilmiştir. Her iki faktörden incelemeye alınan koşullar çaprazlanmış ve 500, 1000 ve 3000 kişiden oluşan, 5, 15 ve 25 maddelik dokuz ayrı veri seti oluşturulmuştur. Oluşturulan veri setlerinden PMTK ve POMTK modellerine göre yetenek kestirimleri yapılmadan önce her iki yaklaşıma göre model veri uyumu incelemesi yapılmıştır. PMTK modellerine göre yapılan incelemelerde, beş maddelik veri setleri için hiçbir PMTK modelinin uyumlu olmadığı bulunmuştur. On beş ve 25 maddeden oluşan veri setlerinde ise, en yüksek uyumun 500 ve 1000 kişilik veri setleri için sağlandığı, 3000 kişilik veri setlerindeki uyumlu madde sayısının daha az olduğu sonucuna ulaşılmıştır. POMTK için yapılan model veri uyumu incelemelerinde ise, beş maddelik veri setlerinden 500 ve 3000 kişilik veri setleri için MHM'nin, 1000 kişilik veri seti için ise, ÇMM'nin uygun olduğu belirlenmiştir. On beş ve 25 maddelik tüm veri setlerinin ise MHM'ye uyum gösterdiği, sonucuna ulaşılmıştır. Test uzunluğu faktörü koşullarına göre oluşturulan veri setlerinden kestirilen yetenekler arasındaki ilişkiler incelendiğinde, şu sonuçlara ulaşılmıştır: testte yer alan madde sayısının beş ve 15 olduğu durumlarda, PMTK kapsamında kestirilen yetenekler ile POMTK kestirimleri ile tutarlı sonuçlar veren tek veri seti 3000 kişiden oluşan veri setidir. Dolayısıyla PMTK'ya dayalı kestirimlerde madde sayısının düşük olduğu durumlarda, örneklem büyüklüğünün 1000'den yüksek olması kestirimleri tutarlı hale getirmektedir. Madde sayısı her iki yaklaşım için de yetenek kestirimleri için oldukça önemli bir faktördür. Hem madde sayısının hem de örneklem büyüklüğünün az olduğu durumlarda iki yaklaşımdan kestirilen yetenekler arasında herhangi bir ilişki olmadığı gözlenmiştir.

Çeşitli Faktörlere Göre Oluşturulan Veri Setlerinin Güvenirliklerinin İncelenmesi

Çalışma kapsamında yanıt aranan son araştırma sorusu incelenen faktörlere göre oluşturulan veri setlerinin güvenirliklerinin PMTK ve POMTK'ya göre nasıl olduğudur. Bu amaçla, PMTK için yapılan güvenirlik analizlerinde test bilgi fonksiyonları ve marjinal güvenirlik katsayıları hesaplanmıştır. POMTK için yapılan güvenirlik analizlerinde ise test düzeyinde bilgi veren dört ayrı katsayı hesaplanmıştır.

PMTK'ya dayalı güvenirlik incelemelerinden elde edilen sonuçlar şu şekilde özetlenebilir:

- Testin verdiği bilgi düzeyi örneklem büyüklüğünden ziyade, test uzunluğundan etkilenmiştir.
- Model veri uyumunun daha iyi sağlandığı 1000 kişilik verilerden elde edilen bilginin ve hesaplanan marjinal güvenirlik katsayılarının tüm madde sayısı koşullarında en yüksek olduğu görülmüştür.
- Beş maddelik veri setlerinin hiçbir örneklem büyüklüğünde yeterli güvenirlik düzeyini sağlamadığı belirlenmiştir.
- Test uzunluğu faktörü içinde incelenen 25 ve 30 maddelik veri setleri için test bilgi (bilgi)fonksiyonlarına göre hesaplanan maksimum bilgi düzeylerinin birbirine çok yakın olduğu bulunmuştur.
- Farklı test uzunluğundan ve örneklem büyüklüklerinden oluşan veri setlerinin tümünün en fazla bilgi verdiği yetenek düzeylerinin sıfır ve negatif değerler etrafında olduğu belirlenmiştir. Bu durum da testte yer alan maddelerin uç yetenek düzeylerini ölçmede yetersiz kaldığını göstermektedir. Testte yer alan maddelerin çoğunluğu, önceki araştırma sorularında incelendiği üzere orta güçlük düzeyindedir, dolayısıyla en fazla bilgiyi de orta yetenek düzeylerinde vermiştir.
- Madde ayırt edicilik parametre değerleri yüksek olan maddelerden oluşan testlerden elde edilen maksimum bilgi düzeyi madde ayırt edicilik parametreleri nispeten düşük olan veri setlerinden elde edilen maksimum bilgi düzeyinden daha yüksek olduğu bulunmuştur.
- Madde güçlük parametreleri yüksek olan maddelerden oluşan testlerden elde edilen bilgi düzeyi diğer veri setlerine göre daha yüksek bulunmuştur.
- İncelenen veri setlerinde yer alan maddelerin ayırt edicilik parametreleri yüksek olsa dahi, testten elde edilen maksimum bilgi düzeyi yetenek ölçeğine orta düzeylere denk gelmektedir. Yalnızca madde güçlüğünün artması, maksimum bilgi düzeyinin pozitif yetenek düzeylerinde olmasını sağlamıştır.

Çalışma kapsamında PMTK ile karşılaştırılan bir diğer yaklaşım olan POMTK ile yapılan güvenirlik incelemeleri için her veri seti için alan yazında önerilen dört ayrı güvenirlik katsayısı hesaplanmıştır, bunlar: Cronbach Alfa, Lambda2, MS ve LCRC katsayılarıdır. Hesaplanan katsayılardan diğer katsayılara göre güvenirliği daha yansız olarak hesapladığı alan yazında belirlenen MS katsayısı temel alınarak yorumlar yapılmıştır (Sijtsma ve Molenaar, 2002; Van Schuur, 2011). Elde edilen değerler Tablo 12'de verilmiştir.

Tablo 12. Farklı faktörlere göre oluşturulan veri setlerinden POMTK yaklaşımına göre kestirilen güvenilirlik katsayıları

Veri setleri	MS	Alfa	Lambda 2	LCRC	Ortalama
30madde500	0.910	0.909	0.910	0.929	0.914
30madde1000	0.918	0.916	0.917	0.931	0.920
30madde3000	0.911	0.910	0.911	0.922	0.913
5madde500	0.626	0.626	0.632	0.698	0.645
5madde1000	0.662	0.652	0.655	0.676	0.661
5madde3000	0.645	0.639	0.643	0.670	0.649
15madde500	0.815	0.810	0.816	0.850	0.822
15madde1000	0.842	0.839	0.841	0.867	0.847
15madde3000	0.839	0.839	0.838	0.857	0.843
25madde500	0.884	0.881	0.884	0.907	0.889
25madde1000	0.901	0.898	0.900	0.916	0.903
25madde3000	0.840	0.841	0.843	0.866	0.847
A-en düşük	0.805	0.803	0.805	0.832	0.811
A-en yüksek	0.880	0.878	0.879	0.900	0.884
B-en kolay	0.842	0.841	0.843	0.866	0.848
B-en zor	0.864	0.861	0.863	0.886	0.868

POMTK ile kestirilen güvenilirlik katsayılarını genel olarak özetlemek gerekirse şu sonuçlara ulaşılabilir:

Madde sayısının artması güvenilirliği artırmıştır.

En yüksek güvenilirlik katsayıları 30 maddelik veri setlerinden kestirilmiştir.

En düşük güvenilirlik katsayıları ise beş maddelik veri setlerinden kestirilmiştir.

Tüm koşullarda en yüksek güvenilirlik katsayıları 1000 kişilik veri setlerinden kestirilmiştir.

Madde ayırt edicilik ve güçlük parametrelerinin değişmesi güvenilirlik katsayılarında büyük değişimlere neden olmamıştır.

Hesaplanan güvenilirlik katsayıları genel olarak birbirine benzer değerler üretse de, en yüksek değer LCRC'den elde edilmiştir. Lambda 2'de diğer katsayılara göre daha yüksektir. Cronbach Alfa ve MS'den kestirilen değerler birbirine daha yakındır. MS'nin daha yansız kestirimler yaptığı POMTK alan yazınında kanıtlanmış olduğu için, LCRC ve Lambda 2 kestirimlerinin gerçek güvenirlüğün üzerinde değerler ortaya çıkardığı sonucuna ulaşılmıştır.

Yukarıda özetlenen bulgulardan yola çıkarak, POMTK ile yüksek düzeyde güvenirlüğün sağlanması için, madde sayısının 15'in üzerinde, örneklem büyüklüğünün ise 1000'e yakın olması gerektiği sonucuna ulaşılmıştır. Beş maddelik veri setlerinin POMTK kapsamında tutarlı sonuçlar vermediği, 15 maddenin tüm örneklemelerde orta düzeyde (0.80'lerde) güvenirlüğe sahip olduğu, ancak 25 madde ve 1000 kişiden oluşan veri setinin güvenirlüğünün 0.90'ın üzerinde olduğu belirlenmiştir. Son olarak elde edilen bulgulardan güvenilirlik değerlerinin örneklem büyüklüğünden çok test uzunluğundan etkileniyor olduğu sonucuna da ulaşılmaktadır. Test uzunluğu arttıkça güvenilirlik artmakta ancak aynı durum örneklem büyüklüğü için sağlanmamaktadır. Örneğin madde sayısı 30 olduğunda, 500 kişilik veri setinde dahi, güvenirlüğün yüksek olduğu bulunmuştur. Bu bulgu Şengül Avşar ve Tavşancıl (2017)'in çalışmasında çok kategorili veriler için elde edilen bulgu ile tutarlıdır ve madde sayısındaki artış, POMTK'da olduğu gibi, POMTK kapsamında da testin güvenilirlik düzeyini artırmıştır.

TARTIŞMA ve SONUÇ

Çalışma kapsamında farklı faktörler ve koşullar gözetilerek oluşturulan veri setlerinin PMTK ve POMTK modellerine uyumu ve kestirilen madde ve yetenek parametreleri arasındaki tutarlılık ile güvenilirlik düzeyleri incelenmiştir. Örneklem büyüklüğü faktörü altında incelenen koşullardan elde edilen sonuçlar genel olarak incelendiğinde, PMTK modellerine göre kestirilen madde parametreleri arasında yüksek ve manidar ilişkiler olduğu belirlenmiş ve madde güçlük parametreleri arasındaki ilişkilerin ayırt edicilik parametreleri arasındaki ilişkilerden daha yüksek olduğu bulunmuştur. Bu durum da madde ayırt edicilik parametrelerinin örneklem büyüklüğünden daha fazla etkilendiği şeklinde yorumlanmış ve alan yazında da benzer bulgulara rastlanmıştır (Meijer, Tenderio ve Wanders, 2014).

POMTK kapsamında yapılan model veri uyumu incelemelerinde hesaplanan Hij katsayıları ile açımlayıcı faktör analizi kapsamında hesaplanan maddelerin faktör yüklerinin birbirine benzer olduğu, her iki yaklaşımda da aynı maddelerin analiz dışında tutulması gerektiği belirlenmiştir. Ayrıca POMTK ile farklı örneklem büyüklüklerinden kestirilen madde parametreleri arasındaki uyumun da yüksek ve manidar olduğu belirlenmiştir. Bu bulgudan yola çıkarak POMTK yaklaşımının küçük örneklerde bile etkili madde parametre kestirimleri sağladığı sonucuna varılmıştır. Bu bulgu da alan yazın da sıklıkla belirtilmekte ve POMTK'nın PMTK'ya avantajı olarak nitelendirilen yönlerden biridir (Junker, 2000; Sijtsma ve Meijer, 2007).

Farklı test uzunluğundan oluşan veri setleri içinde Mokken ölçekleme tekniğine en iyi düzeyde uyum sağlayan veri setinin, 1000 kişilik veri seti olduğu bulunmuştur. Tüm farklı madde sayısı koşullarında 1000 kişilik veri setindeki maddeler için hesaplanan H_i değerlerinin daha yüksek olduğu belirlenmiş dolayısıyla bu veri setleri için hesaplanan H değerleri de daha yüksek olmuştur. Bu bulgu da alan yazında Mokken ölçekleme için gereken en düşük örneklem büyüklüğünü belirlemeyi amaçlayan çalışmada elde edilen bulgular ile paralellik göstermektedir. Maddelere ilişkin ayırt edicilik değerinin veren H_i katsayılarının 0.3'e yakın olması durumunda 1500'den küçük örneklerde yanlış sınıflama oranının yüksek olduğu Straat, van der Ark ve Sijtsma, (2014) tarafından yapılan çalışmada elde edilen bulgulardan biridir. Bu nedenle 500 kişilik veri seti, POMTK için en düşük uyumun elde edildiği veri seti olmuştur.

POMTK ile yapılan ölçekleme sonucu farklı veri setlerinde madde ölçeklenebilirlik katsayıları negatif bulunan 1 ve 12 nolu maddeden, 1 nolu maddenin H_i değerleri 500 ve 3000 kişilik tüm veri setlerinde ve 12 nolu maddenin ise tüm veri setlerinde 0.3'ten düşük olduğu görülmüştür. Dolayısıyla bu maddelerin veri seti dışında tutulmasıyla Mokken ölçeklemeye olan uyumun artacağı sonucuna ulaşılmıştır. Ayrıca bu bulgu 30 maddelik veri setleri için elde edilen bulguyla aynıdır, dolayısıyla madde sayısındaki değişiklik bu maddelere ilişkin hesaplanan H_i değerlerini etkilememiştir. Bu bulgu da alan yazında elde edilen bulgularla benzerlik göstermektedir ve farklı çalışmalarda da belirtildiği üzere test uzunluğunun madde ölçeklenebilirlik katsayılarına olan etkisinin ihmal edilebilir düzeyde olduğu belirlenmiştir (Kuijpers, van der Ark, Croon ve Sijtsma, 2016). Ayrıca tüm veri setleri için, POMTK kapsamında ölçeklemeye uyum sağlamayan madde sayısı, PMTK'ya göre daha düşüktür. Bu durumda alan yazındaki bulgulara benzerdir ve PMTK'nın doğrulayıcı yaklaşımından dolayı ölçeklenemeyen maddelerin POMTK kapsamında ölçeklenmesi mümkün olmaktadır (Meijer ve Baneke, 2004; Sijtsma ve Junker, 2000).

Test uzunluğu koşullarına göre yapılan kestirimlerde, tüm örneklem büyüklükleri ve madde sayısı için POMTK'dan tutarlı yetenek kestirimleri elde edilirken, PMTK ile yalnızca madde sayısının 25 ya da örneklem büyüklüğünün 3000 olduğu durumlarda tutarlı yetenek kestirimleri elde edilmiştir. Her iki yaklaşımdan farklı faktör ve koşullara göre kestirilen yetenek parametreleri arasındaki ilişkilerin manidar ve yüksek olduğu belirlenmiştir. Dolayısıyla yapılacak ölçme uygulamalarında bireylerin yetenek düzeyleri bakımından sıralanması yeterli ise, POMTK kapsamında ölçekleme yapılabileceği sonucuna ulaşılmıştır.

Farklı faktör ve koşullara göre oluşturulan veri setleri için yapılan güvenilirlik incelemesinde, ise PMTK kapsamında beş maddelik veri setlerinden elde edilen test bilgi miktarının 0'a yakın olduğu belirlenmiş ancak test uzunluğunun 15'ten 25'e çıkması testin verdiği maksimum bilgi düzeyini büyük oranda arttırmaktayken, 25'ten 30'a çıkması testin verdiği bilgi düzeyini düşük düzeyde değiştirdiği bulunmuştur. Test uzunluğunun artması PMTK kapsamında güvenilirliği artırırken, örneklem büyüklüğündeki artışın güvenilirlik üzerinde çok etkili olmadığı belirlenmiştir. POMTK ile analiz edilen verilerde en yüksek güvenilirlik 1000 kişilik veri setlerinden elde edilmiş ve madde sayısının artışının PMTK'da olduğu gibi güvenilirlik katsayılarında olumlu yönde değişime neden olduğu belirlenmiştir.

Araştırmada elde edilen bulgulara dayalı olarak, testte yer alan madde sayısının 25'ten az ya da veri setinin geniş olmadığı durumlarda, bireylere ilişkin yetenek kestirimleri sıralama amaçlı kullanılacaksa, POMTK modelleri tercih edilebilir. Araştırmacılar tarafından yalnızca maddelerin ölçeklenmesi isteniyorsa, PMTK modelleriyle incelenen tüm örneklem büyüklüğü ve test uzunluğu koşullarında PMTK modelleriyle yüksek ve manidar ilişkiler gösterdiği çalışma kapsamında belirlenen POMTK modelleri kullanılabilir. Ayrıca psikolojik ölçmeler için değişen madde fonksiyonu belirleme, gelişimi aşamalı olarak takip etme gibi birçok önemli işleve sahip olan değişmez madde sıralaması varsayımı PMTK yaklaşımı ile yapılan araştırmalarda incelenebilir. Böylece maddelerin her yetenek düzeyi için nasıl sıralandığı belirlenebilir.

KAYNAKÇA

- Baker, F. B. and Kim, S.H. (2004). *Item response theory: Parameter estimation techniques* (2nd Edition). New York: Marcel Dekker.
- Christensen, K. and Kreiner, S. (2010). Monte Carlo tests of the Rasch model based on scalability coefficients. *British Journal of Mathematical and Statistical Psychology*, 63, 101-111.
- Dyehouse, M. A. (2009). *A comparison of model-data fit for parametric and nonparametric item response theory models using ordinal-level ratings*. Purdue University.
- Embretson, S. E. and Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.
- Hambleton, R. K. and Swaminathan, H. (1985). *Item response theory. Principles and applications*. Boston: Kluwer.
- Junker, B. (2000). Some topics in nonparametric and parametric IRT. with some thoughts about the future. Carnegie Mellon University: Department of Statistics.
- Koçar, H. (2014). *Madde tepki kuramının farklı uygulamalarından elde edilen parametrelerin ve model uyumlarının örneklem büyüklüğü ve test uzunluğu açısından karşılaştırılması*. Yayınlanmamış doktora tezi. Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü.
- Kujipers, R. E., Van der Ark, L. A., Croon, M. A. and Sijtsma, K. (2016). Bias in point estimates and standard errors of Mokken's scalability coefficients. *Applied Psychological Measurement*. 40 (5), 331-345.
- Loevinger, J. (1947). A systematic approach to the construction and evaluation of tests of ability. *Psychological Monographs*. 61 (A).
- Loevinger, J. (1948). The technic of homogeneous tests compared with some aspects of "scale analysis" and factor analysis. *Psychological Bulletin*. 45 (6). 507-529.
- Meijer, R. R., Sijtsma, K. and Smidt, N. G. (1990). Theoretical and empirical comparison of the Mokken and the Rasch approach to IRT. *Applied Psychological Measurement*. 14. 283-298.
- Meijer, R. R., Sijtsma, K., and Molenaar, I. W. (1995). Reliability estimation for single dichotomous items based on Mokken's IRT model. *Applied Psychological Measurement*, 19 (4), 323-335.
- Meijer, R.R., and Baneke, J.J. (2004). Analyzing psychopathology items: a case for nonparametric item response theory modeling. *Psychological Methods*, 9 (3), 354-368.
- Meijer, R. R., Egberink, I.J.L., Emmons, W.H.M. and 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.
- Meijer, R. R., Tendeiro, J. N., & Wanders, R. B. (2014). The use of nonparametric item response theory to explore data quality. In *Handbook of Item Response Theory Modeling* (s. 103-128). New York: Routledge.
- Mislevy, R. J. & Stocking, M. L. (1989). A consumer's guide to LOJIST and BILOG. *Applied Psychological Measurement*. 13. 57-75.

- Mokken, R. J. and Lewis, C. (1982). A nonparametric approach to the analysis of dichotomous item responses. *Applied Psychological Measurement*, 6 (A), 417-430.
- Molenaar, I.W. veSijtsma, K. (1984).Internalconsistency and realibility in Mokken's nonparametric item response model. *Tijdschriftvoor onderwijsresearch*.9 (5). 257-268.
- Molenaar, I. W. (1991). A weighted Loevinger H coefficient extending Mokken scaling to multicategory items. *KwantitatieveMethoden*. 37. 97- 117.
- Molenaar, I. W. (1997). Nonparametric models for polytomous responses. In W.J. Man der Linden and R.K. Hambleton (Ed.). *Handbook of modern item response theory* (s. 369-380). NewYork: Springer.
- Molenaar, I. W. (2001). Thirty years of nonparametric item response theory. *Applied Psychological Measurement*, 25 (3), 295-299.
- Mokken, R. J. (1971). *A theory and procedure of scale analysis with applications in political research*. New York. Berlin: Walter de Gruyter. Mouton.
- Molenaar, I. W. (1997). Nonparametric models for polytomous responses. In W.J. Man der Linden and R.K. Hambleton (Ed.). *Handbook of modern item response theory* (s. 369-380). NewYork: Springer.
- Sijtsma, K. and Junker, B.W. (2006). Item Response Theory: Past Performance. Present Developments and Future Expectations. *Behaviormetrika*, 1, 75-102.
- Sijtsma, K. and Meijer, R. R. (2007). Nonparametric item response theory. C. R. Rao. & S. Sinharay (Ed.) *Handbook of statistics26: psychometrics*. (s.719-746). Amsterdam: Elsevier.
- Sijtsma, K., and Molenaar, I. W. (2002). *Introduction to nonparametric item response theory*. Newbury Park. CA: Sage.
- Straat, J. H., Van der Ark, L.A. and Sijtsma, K. (2014). Minimum sample size requirements for Mokken scale analysis. *Educational and Psychological Measurement*, 74, 809-822.
- Şengül Avşar, A., & Tavşancıl, E. (2017). Examination of Polytomous Items' Psychometric Properties According to Nonparametric Item Response Theory Models in Different Test Conditions. *Educational Sciences: Theory and Practice*, 17(2), 493-514.
- Van Schuur, W. H. (2011). *Ordinal item response theory: Mokken scale analysis*. Los Angeles: Sage Publications