

Tanımlayıcı ve Açıklayıcı Madde Tepki Modellerinin TIMSS 2007 Türkiye Matematik Verisine Uyarlanması

An Application of Descriptive and Explanatory Item Response Models to TIMSS 2007 Turkey Mathematics Data

Burcu ATAR*

Hacettepe Üniversitesi

Öz

Rasch modeli, bir, iki veya üç parametrelili lojistik modeller gibi madde tepki kuramı (MTK) modelleri, bireylerin ölçülmek istenen örtük özelliklerinin tahmin edilmesinde kullanılan ölçme modelleridirler. MTK modellerinin geleneksel formülasyonları, bireysel farklılıkların açıklanmasına olanak vermemektedirler. Ancak bu modeller genelleştirilmiş doğrusal ve doğrusal olmayan karma modeller çerçevesinde ele alındığı zaman istatistiksel modeller olarak da kullanılabilirler. Genelleştirilmiş doğrusal ve doğrusal olmayan modeller çerçevesinde, tanımlayıcı madde tepki modelleri olarak formüle edilen geleneksel MTK modellerine, bireysel farklılıkları açıklamak üzere birey özelliklerinin ve/veya maddesel farklılıkları açıklamak üzere madde özelliklerinin eklenmesiyle açıklayıcı madde tepki modelleri elde edilir. Bu çalışmada, dört temel tanımlayıcı ve açıklayıcı madde tepki modelinin – Rasch modeli, örtük regresyon Rasch modeli, doğrusal lojistik test modeli ve örtük regresyon doğrusal lojistik test modeli – TIMSS 2007 Türkiye sekizinci sınıf matematik verisi üzerinde uygulanması gösterilmiştir.

Anahtar Sözcükler: Açıklayıcı madde tepki modelleri, örtük regresyon modeli, doğrusal lojistik test modeli, TIMSS 2007.

Abstract

Item response theory (IRT) models such as the Rasch model, one, two, or three parameter logistic models are measurement models that are used to estimate the latent trait of individuals. Traditional formulations of IRT models do not allow explaining individual differences. However, it is possible to use these models as statistical models when they are formulated under the generalized linear and nonlinear mixed models (GLMM and NLMM) framework. Including person properties to explain the differences among person abilities and/or item properties to explain the differences among item difficulties into traditional IRT models that are formulated as descriptive item response models under the generalized linear and nonlinear mixed models framework, explanatory item response models (EIRM) are obtained. In this study, the application of four basic descriptive and explanatory item response models - Rasch model, latent regression Rasch model, linear logistic test model (LLTM), and latent regression LLTM - was illustrated using TIMSS 2007 mathematics data for eight grade Turkish students.

Keywords: Explanatory item response models, latent regression model, linear logistic test model, TIMSS 2007

* Dr. Burcu ATAR, Hacettepe Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, burcua@hacettepe.edu.tr

Summary

Purpose

The purpose of this study is to illustrate the application of four basic descriptive and explanatory item response models - Rasch model, latent regression Rasch model, linear logistic test model (LLTM), and latent regression LLTM - on TIMSS (Trends in International Mathematics and Science Study) 2007 mathematics data

Results

When TIMSS 2007 mathematics data was analyzed under the Rasch model, the variance of mathematics ability was estimated on the person side and the item difficulties were estimated on the item side. The examinee variance was estimated as approximately 1.71 on the logit scale, with a standard error of 0.16, indicating that the differences among examinees are statistically significant at the 0.05 alpha level. The item difficulties were estimated to vary between approximately -1.31 and 3.67 on the logit scale with an average of approximately 0.58 logit, indicating that the probability of correctly answering the average item for an average examinee is lower than 50%.

When the mathematics data was analyzed under the latent regression Rasch model, the effects of gender, positive attitudes toward mathematics, valuing mathematics, and self-confidence in learning mathematics on mathematics ability were estimated on the person side and the item difficulties were estimated on the item side. The effect of only one person property - self-confidence in learning mathematics - on mathematics ability was found to be statistically significant. Self-confidence in learning mathematics explained approximately 23% of the variance of mathematics ability. The residual variance was estimated as approximately 1.32 on the logit scale, with a standard error of 0.13, indicating that the variance that is not explained by person properties in the model is still statistically significant. The item difficulty estimates from the latent regression model were very close to the estimates from the Rasch model.

When the mathematics data was analyzed under the LLTM, the effects of cognitive domain and content domain on item difficulty were estimated on the item side and the examinee variance was estimated on the person side. The effects of both domains on item difficulty were found to be statistically significant. When the cognitive domain was analyzed, it was found that the reasoning items are more difficult than the knowing or applying items. When the content domain was analyzed, it was found that the geometry items are more difficult than the algebra or data analysis and chance items. The residual variance was estimated as approximately 1.21 on the logit scale, with a standard error of 0.12, indicating that the residual variance is statistically significant. It was concluded that the item properties have also effects on the mathematics ability of examinees.

When the mathematics data was analyzed under the latent regression LLTM, the effects of gender, positive attitudes toward mathematics, valuing mathematics, and self-confidence in learning mathematics on mathematics ability were estimated on the person side and the effects of cognitive domain and content domain on item difficulty were estimated on the item side. The estimated person property effects were similar to the estimated effects obtained with the latent regression Rasch model and the estimated item property effects were similar to the estimated effects obtained with the LLTM.

Discussion

The descriptive and explanatory item response models that are analyzed in this study articulate that the Rasch model can be used for both measurement and statistical purposes under the generalized linear mixed models (GLMM) framework. However, the descriptive and explanatory item response models are not restricted to Rasch model only; they can also be applied to other dichotomous item response theory (IRT) models and to polytomous IRT models.

With the flexibility of these models under GLMM framework and the availability of the

programs, it is also possible to add a third level (e.g., school-level) in addition to item-level and person-level.

Conclusion

The Rasch model can be analyzed in the beginning of the study to obtain the measurements of latent ability estimates and the item difficulty estimates but more importantly the estimate of person variance that provides the original variance before the individual differences are explained by person properties and/or by item properties. The value of the original variance can be used to compute the percent of variance explained by the person properties included in the latent regression Rasch model.

If the interest of the study is to explain the differences among persons with respect to the latent ability by person properties and to explain the differences among item difficulties by item properties, it is not necessary to analyze the latent regression Rasch model and the LLTM, instead, the latent regression LLTM can be analyzed.

Giriş

Bu çalışmanın amacı, tanımlayıcı ve açıklayıcı madde tepki modellerinden (descriptive and explanatory item response models) Rasch modeli (Rasch model), örtük regresyon Rasch modeli (latent regression Rasch model), doğrusal lojistik test modeli (linear logistic test model-LLTM) ve örtük regresyon doğrusal lojistik test modelinin (latent regression LLTM) TIMSS 2007 matematik Türkiye verisinin bir örneklemini üzerinde uyarlanmasını göstermektir.

Eğitimde, madde tepki kuramı (MTK) modelleri, öğrencilerin örtük yeteneklerinin ölçülmesi amacıyla geliştirilmiş olan testlerin analizi ve puanlanması için kullanılırlar. İki kategorili (dichotomous) veya çok kategorili (polytomous) puanlanan verilerin incelenmesi için çeşitli madde tepki kuramı modelleri bulunmaktadır. İki kategorili puanlanan veriler için örnek olarak, Rasch modeli, bir, iki ve üç parametrelili lojistik (1PL, 2PL, 3 PL) modeller (Rasch 1960; Lord ve Novick, 1968; Embretson ve Reise, 2000), çok kategorili puanlanan veriler için ise kısmi puan modeli (partial credit model; Masters, 1982), aşamalı tepki modeli (graded response model; Samejima, 1969) ve derecelendirme ölçeği modeli (rating scale model; Andrich, 1978a, 1978b) örnek olarak gösterilebilirler. Bu geleneksel MTK modelleri, bireylerin performanslarını yetenek veya başarı puanları cinsinden, madde niteliklerini ise güçlük ve/veya ayrıcalık cinsinden tanımlamak için kullanılmaları sebebiyle “tanımlayıcı (descriptive)” ölçme modelleri olarak ele alınırlar (De Boeck ve Wilson, 2004).

MTK modelleri, genelleştirilmiş doğrusal ve doğrusal olmayan karma modeller (genelleştirilmiş doğrusal karma modeller: Breslow ve Clayton, 1993; McCulloch ve Searle, 2001, doğrusal olmayan karma modeller: Davidian ve Giltinan, 1995) çerçevesinde ele alındığında, birey ve/veya madde yordayıcılarının modele eklenmesiyle, bireylerin performansları ve/veya maddelerin güçlükleri arasında gözlenen farklılıkların açıklanması mümkün olmaktadır (De Boeck ve Wilson, 2004; Skrondal ve Rabe-Hesketh, 2004; Briggs, 2008). Böylece, MTK modelleri tanımlayıcı ölçme modelleri olarak kullanılmalarının yanı sıra “açıklayıcı (explanatory)” istatistiksel modeller olarak da kullanılabilirler. De Boeck ve Wilson (2004) tarafından bu modeller, *açıklayıcı madde tepki modelleri (explanatory item response models, EIRM)* olarak adlandırılmışlardır.

Genelleştirilmiş Doğrusal Karma Modeller

Bu çalışmada ele alınan tanımlayıcı ve açıklayıcı madde tepki kuramı modelleri, genelleştirilmiş doğrusal karma modellerin (generalized linear mixed models) özel formlarıdır. Gözlemlerin sürekli olmadığı durumlarda, gözlemlerin, bu gözlemleri doğrusal olarak açıklayacak sabit ve rastlantısal yordayıcılara bir bağlantı fonksiyonu ile ilişkilendirildiği modeller, genelleştirilmiş doğrusal karma modellerdir. Bu modeller, bağımlı değişken olarak sürekli değişkenlerin ele alındığı doğrusal modellerin süreksiz değişkenlere genişletildiği ve

yordayıcı değişkenlerin sabit etkilerinin (fixed effects) yanı sıra rastlantısal etkilerinin (random effects) de ele alındığı modelleri kapsadığı için bu adı almışlardır.

Genelleştirilmiş doğrusal karma modeller, bağımlı değişkenin dağılım fonksiyonunun tanımlandığı *dağılım ailesi* (distributional family), dağılımın ortalama değerinin doğrusal bileşene bağlandığı *bağlantı fonksiyonu* (link function) ve yordayıcı değişkenlerin doğrusal bir fonksiyon olarak tanımlandığı *doğrusal bileşen* (linear component) olmak üzere üç parçadan oluşurlar (McCulloch & Searle, 2001; De Boeck & Wilson, 2004; Rabe-Hesketh, Pickles, & Skrondal, 2004; Skrondal, & Rabe-Hesketh, 2004).

Genelleştirilmiş doğrusal karma modeller, MTK modellerinin iki seviyeli modeller olarak formüle edilmesine olanak tanır. Bu durumda, en düşük seviyedeki (seviye 1) analiz birimi bir testteki maddelerdir, bir üst seviyedeki (seviye 2) analiz birimi ise bu testteki maddelere cevap veren bireylerdir (De Boeck ve Wilson, 2004). MTK modellerinin iki seviyeli formülasyonunda, belli bir örtük yeteneği (latent ability) kestirmek için hazırlanan bir ölçme aracındaki herhangi bir maddeden belli bir puanı alma olasılığı, dağılım ailesi, bağlantı fonksiyonu ve doğrusal bileşen yoluyla belirlenir (De Boeck ve Wilson, 2004; Rabe-Hesketh, Pickles, ve Skrondal, 2004; Skrondal, ve Rabe-Hesketh, 2004).

Bir testteki i maddesini cevaplayan p bireyi için madde puanının (bağımlı değişken, Y_{pi}) 1 veya 0 değerini aldığı bir örnekte, dağılım ailesi olarak Bernoulli dağılımı kullanılır. Y_{pi} bağımlı değişkeninin Bernoulli dağılım fonksiyonu, dağılımın beklenen değeri (ortalama değeri), π_{pi} ile tanımlanır. ($Y_{pi} \sim \text{Bernoulli}(\pi_{pi})$). π_{pi} , p bireyinin i maddesini doğru cevaplama olasılığı ($Y_{pi} = 1$) olarak ifade edilir.

1 ve 0 değerlerinden oluşan gözlemlerin ortalama değerleri, π_{pi} , normal ojiv modeller (normal-ogive models) için probit, lojistik modeller (logistic models) için ise lojit bağlantı fonksiyonu ile doğrusal bileşene bağlanır.

Genelleştirilmiş doğrusal karma modellerin doğrusal bileşen parçası genel olarak aşağıdaki eşitlik ile formüle edilir:

$$\eta_{pi} = \sum_{k=0}^K \beta_k X_{ik} + \sum_{j=0}^J \theta_{pj} Z_{ij} \quad (1)$$

Bu eşitlikteki,

X_{ik} , i maddesinin k yordayıcı değişkeni için aldığı değerdir,

β_k , k yordayıcı değişkenin etkisi olup $k=0$ için kesişim, $k=1, \dots, K$ için ise eğim katsayısıdır ve bu katsayıların değeri bireyden bireye değişmediği için sabit etkiler olarak adlandırılırlar,

Z_{ij} , i maddesinin j yordayıcı değişkeni için aldığı değerdir,

θ_{pj} , p bireyi için j yordayıcı değişkenin etkisi olup $j=0$ için p bireyinin kesişim değerinden (β_0) sapmasıdır, $j=1, \dots, J$ için ise p bireyinin ilgili eğim değerinden (β_1, \dots, β_j) sapmasıdır ve bu etkilerin değeri bireyden bireye değiştiği için rastlantısal etkiler olarak adlandırılırlar. θ_p 'nin normal dağıldığı, ortalamasının sıfır, varyansının ise σ_p^2 olduğu varsayılır.

Bu çalışmada ele alınan Rasch modeli, örtük regresyon Rasch modeli, doğrusal lojistik test modeli ve örtük regresyon doğrusal lojistik test modeli, sadece kesişim katsayısının rastlantısal olarak dağıldığı lojistik rastlantısal-kesişim modelleridirler (logistic random-intercepts models). Lojistik rastlantısal-kesişim modellerinin doğrusal bileşen parçası genel olarak aşağıdaki eşitlik ile formüle edilir:

$$\eta_{pi} = \sum_{k=0}^K \beta_k X_{ik} + \theta_{p0} Z_{i0} \quad (2)$$

Rasch Modeli

Rasch modeli, sadece madde ve birey parametrelerinin kestirildiği, gerek madde kısmında, gerekse birey kısmında kestirilen parametre değerleri arasındaki farklılıkları açıklamak üzere herhangi bir yordayıcı değişkenin bulunmadığı tanımlayıcı bir modeldir.

Eşitlik 2'deki X_{ik} , madde yordayıcı değişkeni, i maddesini belirlemek üzere madde belirleyicisi (item indicator) olarak kullanıldığı zaman ($i = k$ ise $X_{ik} = 1$, $i \neq k$ ise $X_{ik} = 0$), eşitlikteki $\sum_k \beta_k X_{ik}$ terimi madde parametresine denk gelir ve β_i ile gösterilir (De Boeck ve Wilson, 2004). $-\beta_i$, i maddesi için madde güçlük parametresidir. Bu modelde, madde sayısı kadar madde belirleyicisi vardır. Modelin tanımlanabilmesi için, bu modelden kesişim katsayısı, β_0 , çıkarılmıştır.

Eşitlik 2'deki $\theta_p Z_{i0}$ terimi birey parametresine denk gelir ve θ_p ile gösterilir (De Boeck ve Wilson, 2004). θ_p , p bireyi için yetenek düzeyidir. Böylece eşitlik 2 aşağıdaki şekli alır:

$$\eta_{pi} = \beta_i + \theta_p \quad (3)$$

Bu eşitlikteki i maddesini cevaplayan p bireyi için doğrusal bileşen, η_{pi} , logit bağlantı fonksiyonundan dolayı p bireyinin i maddesini doğru cevaplama olasılığının ($Y_{pi} = 1$) logit fonksiyonuna eşit olup $\eta_{pi} = \text{logit}(\pi_{pi}) = \ln\left(\frac{\pi_{pi}}{1-\pi_{pi}}\right)$ şeklinde tanımlanır ki bu eşitlikteki $\left(\frac{\pi_{pi}}{1-\pi_{pi}}\right)$ terimi odds oranı olarak adlandırılır. Eşitlik 3'te, eşitliğin her iki tarafının da üsseli alındığında odds oranını ve olasılığı hesaplamak mümkündür:

$$\exp(\eta_{pi}) = \exp\left(\ln\left(\frac{\pi_{pi}}{1-\pi_{pi}}\right)\right) = \exp(\beta_i + \theta_p) \quad (4)$$

Bu durumda,

$$\left(\frac{\pi_{pi}}{1-\pi_{pi}}\right) = \exp(\beta_i + \theta_p) \quad (5)$$

$$\pi_{pi} = \frac{\exp(\beta_i + \theta_p)}{1 + \exp(\beta_i + \theta_p)} \quad (6)$$

Örtük Regresyon Rasch Modeli

Örtük regresyon Rasch modeli, bireylerin yetenekleri arasındaki farklılıkları açıklamak üzere birey kısmında yordayıcı değişkenler olarak birey özelliklerinin eklendiği bir modeldir. Bu durumda eşitlik 3'teki θ_p birey parametresi bağımlı değişken olarak ele alınarak bu bağımlı değişkeni açıklayacak doğrusal regresyon eşitliği kurulur:

$$\theta_p = \sum_{j=1}^J \vartheta_j Z_{pj} + \varepsilon_p \quad (7)$$

Bu eşitlikteki,

Z_{pj} , p bireyi için j yordayıcı değişkeninin aldığı değerdir,

ϑ_j , j yordayıcı değişkenin sabit etkisidir,

ε_p , birey özelliklerinin etkisi hesaba alındığı zaman geriye kalan birey etkisi olup Z_{p0} sabit değişkeninin rastlantısal etkisi olarak düşünülebilir. ε_p 'nin normal dağıldığı, ortalamasının sıfır, varyansının ise σ_ε^2 olduğu varsayılır.

Eşitlik 3'teki θ_p , eşitlik 7'deki doğrusal regresyon eşitliği ile tanımlandığında, eşitlik 3 aşağıdaki formunu alır:

$$\eta_{pi} = \beta_i + \sum_{j=1}^J \vartheta_j Z_{pj} + \varepsilon_p \quad (8)$$

Doğrusal Lojistik Test Modeli

Doğrusal lojistik test modeli, maddelerin güçlükleri arasındaki farklılıkları açıklamak üzere madde kısmında yordayıcı değişkenler olarak madde özelliklerinin eklendiği bir modeldir. Bu durumda eşitlik 3'teki β_i madde parametresi bağımlı değişken olarak ele alınarak bu bağımlı değişkeni açıklayacak doğrusal bir fonksiyon kurulur:

$$\beta_i = \sum_{k=0}^K \beta_k X_{ik} \quad (9)$$

Bu eşitlikteki,

X_{ik} , i maddesinin k yordayıcı değişkeni için aldığı değerdir,

β_k , k yordayıcı değişkenin etkisidir.

Bu modeldeki madde yordayıcıları, β_i madde parametresini mükemmel olarak açıklamadığı sürece $-\beta_i$ değeri, madde güçlük parametresine eşit olmayacaktır. Eşitlik 3'teki β_i , eşitlik 9'daki doğrusal fonksiyon ile tanımlandığında, eşitlik 3 aşağıdaki formunu alır:

$$\eta_{pi} = \sum_{k=0}^K \beta_k X_{ik} + \theta_p \quad (10)$$

Doğrusal lojistik test modeli, eşitlik 9'a hata teriminin, ε_i , eklenmesiyle daha esnek bir forma dönüşebilir. Ancak GLLAMM yazılımı bu modele hata terimi eklenmesine izin vermemektedir.

Örtük Regresyon Doğrusal Lojistik Test Modeli

Örtük regresyon doğrusal lojistik test modeli, bireylerin yetenekleri arasındaki farklılıkları açıklamak üzere birey kısmında yordayıcı değişkenler olarak birey özelliklerinin eklendiği, maddelerin güçlükleri arasındaki farklılıkları açıklamak üzere madde kısmında yordayıcı değişkenler olarak madde özelliklerinin eklendiği bir modeldir. Bu durumda eşitlik 3 aşağıdaki şekilde düzenlenir:

$$\eta_{pi} = \sum_{k=0}^K \beta_k X_{ik} + \sum_{j=1}^J \vartheta_j Z_{pj} + \varepsilon_p \quad (11)$$

Yöntem

Bu çalışmada, tanımlayıcı ve açıklayıcı madde tepki modellerinin uygulanmasını göstermek üzere, TIMSS (Trends in International Mathematics and Science Study) 2007 Türkiye sekizinci sınıf matematik verisinin bir bölümü kullanılmıştır.

TIMSS, Uluslararası Eğitim Başarılarını Değerlendirme Kuruluşu (International Association for the Evaluation of Educational Achievement, IEA) tarafından ilki 1995 yılında olmak üzere dört yılda bir gerçekleştirilen uluslararası bir çalışmadır. Bu çalışmada, iki aşamalı tabakalı örnekleme yöntemi ile seçilen dördüncü ve sekizinci sınıftaki öğrencilere matematik ve fen başarı testleri uygulanır. 2007 yılındaki uygulamada, matematik başarı testi her biri iki bloktan oluşan yedi kitapçık olarak hazırlanmıştır. Her bir öğrenci yalnız bir kitapçığı cevaplar. Başarı testlerinin yanında öğrencilere uygulanan öğrenci anketleri, öğretmenlere uygulanan öğretmen anketleri ile müdürlere uygulanan okul anketleri de bulunmaktadır.

Bu çalışmanın kapsamında, analizlerin ve uygulamanın sadeliği için, TIMSS 2007 matematik başarı testi için hazırlanan kitapçıklardan kitapçık 6'yı tamamlayan 319 öğrencinin verisi kullanılmıştır. Bu öğrencilerden %51'i erkek, %49'u ise kızdır. Kitapçık 6, 31 tane iki kategorili puanlanan, 1 tane de üç kategorili puanlanan toplam 32 madde içerir. Üç kategorili puanlanan madde analiz dışı bırakılmıştır.

Rasch modelinde, öğrencilerin matematik başarı testindeki (kitapçık 6) 31 maddeden aldığı puanlar kullanılmıştır. Bu 31 madde iki kategorili olarak puanlandığı için öğrenciler her bir maddeden doğru cevap için 1, yanlış veya boş cevap için 0 puan almışlardır.

Örtük regresyon Rasch ve örtük regresyon doğrusal lojistik test modellerinde, öğrencilerin matematik başarılarında gözlenen farklılıkları açıklamak üzere birey özellikleri ile ilgili dört yordayıcı değişken kullanılmıştır: Cinsiyet, matematiğe karşı olumlu tutum, matematiğe önem verme ve matematik öğrenmede özgüven. Bu değişkenler ile ilgili betimsel istatistikler Tablo 1’de verilmiştir.

Cinsiyet (CİNSİYET): Cinsiyet değişkeni iki kategorili bir değişken olup erkek öğrenciler için “1”, kız öğrenciler için ise “0” olarak kodlanmıştır.

Matematiğe Karşı Olumlu Tutum (TUTUM): Matematiğe karşı olumlu tutum değişkeni sürekli bir değişkendir. Öğrencilerin, öğrenci anketindeki “matematik öğrenmekten zevk alıyorum”, “matematik sıkıcıdır”, “matematiği seviyorum” ve “daha fazla matematik almak istiyorum” ifadelerinden aldığı puanların toplanmasıyla oluşturulmuştur. Öğrenciler her bir ifade için “kesinlikle katılıyorum”, “katılıyorum”, “katılmıyorum” ve “kesinlikle katılmıyorum” şeklinde cevap vermişlerdir. “Kesinlikle katılıyorum” cevabı için 4 puan, “katılıyorum” cevabı için 3 puan, “katılmıyorum” cevabı için 2 puan ve “kesinlikle katılmıyorum” cevabı için 1 puan almışlardır. Bu puanlar sadece “matematik sıkıcıdır” ifadesi için ters olarak kodlanmıştır. Böylece bu değişkenin aldığı en düşük değer 4, en yüksek değer ise 16’dır.

Matematiğe Önem Verme (ÖNEM): Matematiğe önem verme değişkeni sürekli bir değişkendir. Öğrencilerin, öğrenci anketindeki “matematik öğrenmenin günlük yaşamımda bana yardımcı olacağını düşünüyorum”, “okuldaki diğer dersleri öğrenmek için matematiğe ihtiyacım var”, “istediğim üniversiteye girebilmek için matematikte başarılı olmalıyım” ve “istediğim işe girebilmek için matematikte başarılı olmalıyım” ifadelerinden aldığı puanların toplanmasıyla oluşturulmuştur. Öğrenciler her bir ifade için “kesinlikle katılıyorum”, “katılıyorum”, “katılmıyorum” ve “kesinlikle katılmıyorum” şeklinde cevap vermişlerdir. Bu değişkenin aldığı en düşük değer 4, en yüksek değer ise 16’dır.

Matematik Öğrenmede Özgüven (ÖZGÜVEN): Matematik öğrenmede özgüven değişkeni sürekli bir değişkendir. Öğrencilerin, öğrenci anketindeki “matematikte genellikle başarılıyım”, “matematik benim için sınıf arkadaşlarıma göre daha zordur”, “matematikte iyi değilim” ve “matematik konularını çabuk öğrenirim” ifadelerinden aldığı puanların toplanmasıyla oluşturulmuştur. Öğrenciler her bir ifade için “kesinlikle katılıyorum”, “katılıyorum”, “katılmıyorum” ve “kesinlikle katılmıyorum” şeklinde cevap vermişlerdir. Puanlar “matematik benim için sınıf arkadaşlarıma göre daha zordur” ve “matematikte iyi değilim” ifadeleri için ters olarak kodlanmıştır. Bu değişkenin aldığı en düşük değer 4, en yüksek değer ise 16’dır.

Doğrusal lojistik test ve örtük regresyon doğrusal lojistik test modellerinde, maddelerin güçlükleri arasında gözlenen farklılıkları açıklamak üzere iki madde özelliği kullanılmıştır: bilişsel alan ve konu alanı.

Bilişsel alan, bilgi, uygulama ve akıl yürütme olmak üzere üç kategoriden oluşan bir değişkendir.

Bilgi: Bu alan ile ilgili maddeler öğrencilerin matematik ile ilgili bilmesi gereken olguları, kavramları ve yöntemleri kapsar. Öğrencilerin matematiği uygulama ve matematikle ilgili durumlarda akıl yürütme becerileri matematiksel bilgilerine bağlıdır.

Uygulama: Bu alan ile ilgili maddeler, öğrencilerin matematikle ilgili problemleri çözmek veya sorulara cevap vermek için bilgilerini uygulayabilme yeteneklerine odaklanır.

Akıl Yürütme: Bu alan ile ilgili maddeler, öğrencilerin alışılmış problemlerin dışındaki problemleri mantıksal ve sistematik düşünme yoluyla çözebilme yeteneklerine odaklanır.

Bilişsel alan değişkeninin kategorilerini tanımlamak için akıl yürütme kategorisi referans alınarak modele iki yordayıcı değişken eklenmiştir: bilgi – akıl yürütme (BİLGİ) ve uygulama –

akıl yürütme (UYGULA). Bu durumda, bilişsel alan ile ilgili birinci yordayıcı değişken (BİLGİ) için bilgi gerektiren maddeler "1", uygulama gerektiren maddeler "0" ve akıl yürütme gerektiren maddeler "0" olarak, ikinci yordayıcı değişken (UYGULA) için ise bilgi gerektiren maddeler "0", uygulama gerektiren maddeler "1" ve akıl yürütme gerektiren maddeler "0" olarak kodlanmıştır.

Konu alanı, sayılar, cebir, veri analizi ve olasılık ve geometri olmak üzere dört kategoriden oluşan bir değişkendir.

Sayılar: Bu alan ile ilgili maddeler, tam sayılar, kesirler ve ondalık sayılar, sayı basamakları, üstler, tahmini hesaplama ve çok yakın tahmin ve oran-orantı hesaplarını kapsar.

Cebir: Bu alan ile ilgili maddeler, sayı dizileri, sayısal durumların gösterilmesi, basit doğrusal denklemleri çözme, ifadelerle işlemler, ilişkiler ve fonksiyonların gösterilmesini kapsar.

Veri Analizi ve Olasılık: Bu alan ile ilgili maddeler, çizelgeleri, grafikleri ve tabloları gösterme ve yorumlama; temel istatistiksel hesaplamalar, olasılık ve basit sayısal olasılığı kapsar.

Geometri: Bu alan ile ilgili maddeler, noktalar, çizgiler, düzlemler, açılar, görselleştirme, üçgenler, çokgenler, daireler, dönüşümler, simetri, eşitlik, benzerlik ve bazı temel çizimleri kapsar.

Konu alanı değişkenin kategorilerini tanımlamak için geometri kategorisi referans alınarak modele üç yordayıcı değişken eklenmiştir: sayılar – geometri (SAYI), cebir – geometri (CEBİR) ve veri analizi ve olasılık – geometri (VERİ). Bu durumda, konu alanı ile ilgili birinci yordayıcı değişken (SAYI) için sayılar ile ilgili maddeler "1", cebir ile ilgili maddeler "0", veri analizi ve olasılık ile ilgili maddeler "0" ve geometri ile ilgili maddeler "0" olarak, ikinci yordayıcı değişken (CEBİR) için sayılar ile ilgili maddeler "0", cebir ile ilgili maddeler "1", veri analizi ve olasılık ile ilgili maddeler "0" ve geometri ile ilgili maddeler "0" olarak, üçüncü yordayıcı değişken (VERİ) için ise sayılar ile ilgili maddeler "0", cebir ile ilgili maddeler "0", veri analizi ve olasılık ile ilgili maddeler "1" ve geometri ile ilgili maddeler "0" olarak kodlanmıştır.

Tablo 1.

Öğrenci Özellikleri ile İlgili Değişkenlerin Betimsel İstatistikleri

Öğrenci Özellikleri	Ortalama	Standart Sapma	En Düşük Değer	En Yüksek Değer
TUTUM	12.88	2.79	4	16
ÖNEM	14.06	2.07	4	16
ÖZGÜVEN	10.71	3.15	4	16

Verinin analize hazır bir hale getirilmesi için SAS programından yararlanılmıştır. Analizler için ise STATA (StataCorp., 2003) üzerinden çalışan GLLAMM (Rabe-Hesketh, Skrondal, ve Pickles, 2004) yazılımı kullanılmıştır. Açıklayıcı madde tepki modelleri yaklaşımı altında GLLAMM yazılımı ile gerçekleştirilen Rasch modeli analizinden elde edilen madde güçlük parametre tahminlerini karşılaştırmak üzere MULTILOG programı ile geleneksel Rasch modeli analizi yapılmıştır.

Bulgular

Rasch Modeli

Gruptaki öğrencilerin matematik başarılarının varyansı lojit ölçeğinde yaklaşık 1.71 olarak tahmin edilmiştir; bu tahminin standart hatası 0.16'dır. Bu durumda, öğrencilerin matematik başarılarında gözlenen farklılıklar 0.05 alfa düzeyinde istatistiksel olarak anlamlıdır (t -değeri = $1.71/0.16 = 10.69$, $t_{(0.05, 218)} = 1.96$; $p < 0.001$). Lojit ölçeğinde tahmin edilen varyans değerinin karekökü alındığında ($\sqrt{1.71}$) elde edilen 1.31 değeri öğrencilerin matematik başarılarının standart sapmasını verir. Bu değer üsseli alındığında ($\exp(1.31)$) elde edilen 3.71 değeri ise odds oranını verir (Odds oranı, testteki herhangi bir maddenin doğru cevaplanma olasılığının yanlış cevaplanma olasılığına oranı olup doğru cevaplanma olasılığının 0.50 olması durumunda

1'e eşittir. Odds oranının 3.71'e eşit olduğu durumda, maddenin doğru cevaplanma olasılığı yaklaşık 0.79'dur). O halde, matematik başarı puanları arasında 1 standart sapma (1.31 lojit) olan iki öğrenciden, puanı diğerine göre 1 standart sapma fazla olan öğrencinin odds oranı, diğer öğrenciye göre yaklaşık 3.71 kat daha fazladır. Diğer bir ifade ile testteki herhangi bir maddeyi doğru cevaplama olasılığı 0.50 olan bir öğrenciye göre, matematik başarı puanı 1 standart sapma yüksek olan bir öğrencinin bu maddeyi doğru cevaplama olasılığı yaklaşık 0.79'dur.

MULTILOG programı ve GLLAMM yazılımı ile elde edilen madde güçlük parametre tahminleri ve bu tahminlerin standart hataları Tablo 2'de verilmiştir. GLLAMM yazılımı ile elde edilen tahminler, MULTILOG programı ile elde edilen tahminlere çok yakındır. MULTILOG programı ile elde edilen madde güçlük parametre tahminleri lojit ölçeğinde yaklaşık -1.28 ile 3.53 değerleri arasında değişmekte olup ortalama madde güçlüğü yaklaşık 0.57'dir. GLLAMM yazılımı ile elde edilen madde güçlük parametre tahminleri ise lojit ölçeğinde yaklaşık -1.31 ile 3.67 değerleri arasında değişmekte olup ortalama madde güçlüğü yaklaşık 0.58'dir. Eşitlik 6'da matematik başarısı açısından ortalama bir öğrenciyi temsil etmek üzere θ_p yerine 0 değeri, ortalama madde güçlüğüne sahip bir maddeyi temsil etmek üzere de β_i yerine 0.58 değeri yerleştirildiğinde elde edilen 0.36 değeri şu şekilde yorumlanabilir: Matematik başarısı açısından ortalama bir öğrencinin, ortalama madde güçlüğüne sahip bir maddeyi doğru olarak cevaplayabilme olasılığı yaklaşık 0.36'dır.

Tablo 2.

MULTILOG ve GLLAMM ile Madde Güçlük Parametre Tahminleri ve Standart Hataları

	MULTILOG	GLLAMM	
	Rasch Modeli Tahmin (Standart Hata)	Rasch Modeli Tahmin (Standart Hata)	Ortük Regresyon Rasch Modeli Tahmin (Standart Hata)
Madde 1	-0.41 (0.17)	-0.43 (0.15)	-0.42 (0.53)
Madde 2	3.53 (0.35)	3.67 (0.28)	3.69 (0.59)
Madde 3	-0.41 (0.18)	-0.43 (0.15)	-0.42 (0.53)
Madde 4	-1.01 (0.19)	-1.05 (0.15)	-1.03 (0.53)
Madde 5	1.49 (0.20)	1.54 (0.17)	1.55 (0.54)
Madde 6	1.38 (0.22)	1.42 (0.17)	1.44 (0.54)
Madde 7	0.65 (0.19)	0.67 (0.15)	0.68 (0.53)
Madde 8	-0.54 (0.16)	-0.56 (0.15)	-0.55 (0.53)
Madde 9	-0.40 (0.16)	-0.41 (0.15)	-0.40 (0.53)
Madde 10	1.27 (0.18)	1.31 (0.17)	1.33 (0.54)
Madde 11	0.33 (0.18)	0.33 (0.15)	0.34 (0.53)
Madde 12	0.16 (0.16)	0.16 (0.15)	0.18 (0.53)
Madde 13	0.08 (0.15)	0.08 (0.15)	0.09 (0.53)
Madde 14	1.27 (0.20)	1.31 (0.17)	1.33 (0.54)
Madde 15	-0.80 (0.19)	-0.83 (0.15)	-0.82 (0.53)
Madde 16	0.81 (0.19)	0.83 (0.16)	0.84 (0.53)
Madde 17	-0.59 (0.19)	-0.61 (0.15)	-0.60 (0.53)
Madde 18	0.24 (0.18)	0.25 (0.15)	0.26 (0.53)
Madde 19	0.63 (0.20)	0.65 (0.15)	0.66 (0.53)
Madde 20	-0.40 (0.16)	-0.41 (0.15)	-0.40 (0.53)
Madde 21	0.97 (0.18)	1.00 (0.16)	1.02 (0.54)
Madde 22	1.27 (0.22)	1.31 (0.17)	1.33 (0.54)
Madde 23	1.67 (0.23)	1.74 (0.18)	1.75 (0.54)
Madde 24	1.40 (0.22)	1.45 (0.17)	1.46 (0.54)
Madde 25	2.32 (0.28)	2.41 (0.20)	2.43 (0.55)
Madde 26	1.38 (0.20)	1.40 (0.17)	1.42 (0.54)
Madde 27	1.01 (0.20)	1.04 (0.16)	1.06 (0.54)
Madde 28	1.07 (0.18)	1.10 (0.16)	1.12 (0.54)
Madde 29	-0.66 (0.18)	-0.68 (0.15)	-0.67 (0.53)
Madde 30	1.11 (0.21)	1.15 (0.16)	1.16 (0.54)
Madde 31	-1.28 (0.19)	-1.31 (0.16)	-1.29 (0.53)

Örtük Regresyon Rasch Modeli

Cinsiyet (CİNSİYET), matematiğe karşı olumlu tutum (TUTUM), matematiğe önem verme (ÖNEM) ve matematik öğrenmede özgüven (ÖZGÜVEN) yordayıcılarının öğrencilerin matematik başarıları üzerindeki etkilerinin tahminleri ile bu tahminlerin standart hataları, ilgili t -değerleri ve p -değerleri Tablo 3'te verilmiştir. t -değerleri ve p -değerleri incelendiğinde, cinsiyet, matematiğe karşı olumlu tutum, matematiğe önem verme ve matematik öğrenmede özgüven yordayıcılarından sadece öğrencilerin matematik öğrenmede kendilerine olan güvenlerinin matematik başarıları üzerinde 0.05 alfa düzeyinde istatistiksel olarak anlamlı olduğu görülmektedir.

Tablo 3.

Örtük Regresyon Rasch Modeli İçin Öğrenci Parametre Tahminleri

Öğrenci Özellikleri	Parametre	Tahmin	Standart Hata	df	t -değeri (p -değeri)
CİNSİYET	β_1	-0.09	0.14	318	-0.64 (0.52)
TUTUM	β_2	-0.01	0.03	318	-0.33 (0.74)
ÖNEM	β_3	0.02	0.04	318	0.50 (0.62)
ÖZGÜVEN	β_4	0.20	0.03	318	6.67 (<0.001)

Öğrencilerin matematik öğrenmede kendilerine olan güvenlerinin matematik başarıları üzerindeki etkisi lojit ölçeğinde 0.03 standart hata ile yaklaşık 0.16 olarak tahmin edilmiştir. 0.05 alfa düzeyinde istatistiksel olarak anlamlı bulunan bu değer, matematik öğrenmede özgüven puanları arasında bir birimlik fark bulunan iki öğrencinin matematik başarı puanları arasında tahmini 0.16 lojitlik bir fark olacağı anlamına gelmektedir. Lojit ölçeğindeki 0.16 değerinin üsseli alındığında ($\exp(0.16)$) elde edilen 1.17 değeri odds oranını verir. O halde, matematik başarı puanları arasında 0.16 lojit olan iki öğrenciden, puanı diğerine göre 0.16 lojit fazla olan öğrencinin odds oranı, diğer öğrenciye göre yaklaşık 1.17 kat daha fazladır. Diğer bir ifade ile testteki herhangi bir maddeyi doğru cevaplama olasılığı 0.50 olan bir öğrenciye göre, matematik başarı puanı 0.16 lojit yüksek olan bir öğrencinin bu maddeyi doğru cevaplama olasılığı yaklaşık 0.54'tür. Öğrencilerin matematik öğrenmede kendilerine olan güvenlerinin matematik başarıları üzerindeki etkisi standart sapma açısından da yorumlanabilir. 0.03 olarak bulunan standart hata, öğrenci sayısının karekökü ($\sqrt{319}$) ile çarpıldığında bulunan 0.54 değeri standart sapmayı vermektedir. Lojit ölçeğindeki 0.54 değerinin üsseli alındığında ($\exp(0.54)$) elde edilen 1.72 değeri odds oranını verir. O halde, matematik öğrenmede özgüven puanları arasında bir standart sapma fark bulunan iki öğrenciden, puanı diğerine göre 1 standart sapma fazla olan öğrencinin odds oranı, diğer öğrenciye göre yaklaşık 1.72 kat daha fazladır. Diğer bir ifade ile testteki herhangi bir maddeyi doğru cevaplama olasılığı 0.50 olan bir öğrenciye göre, matematik öğrenmede kendine güven puanı 1 standart sapma yüksek olan bir öğrencinin bu maddeyi doğru cevaplama olasılığı yaklaşık 0.63'tür.

Öğrencilerin matematik başarılarında görülen farklılıkların bir kısmı cinsiyet, matematiğe karşı olumlu tutum, matematiğe önem verme ve matematik öğrenmede özgüven değişkenleri ile açıklandıktan sonra bireylerin matematik başarılarının geriye kalan varyansı beklenildiği gibi Rasch modelinde tahmin edilen değerden (1.71 lojit) daha düşük olup lojit ölçeğinde yaklaşık 1.32 olarak tahmin edilmiştir; bu tahminin standart hatası 0.13'tür. Varyansın açıklanmayan bu kısmı, 0.05 alfa düzeyinde hâla istatistiksel olarak anlamlıdır ($p < 0.001$).

Matematik öğrenmede özgüven değişkeni tarafından açıklanan varyans, bu değişkenin varyansının değişkenin etkisinin karesinin çarpımı ($3.15^2 \times 0.20^2$) ile elde edilen değerdir. Hesaplanan bu değer yaklaşık 0.40 olup toplam varyansa (1.71) bölüldüğünde elde edilen 0.23 değeri, varyansın yaklaşık %23'lük bir kısmının matematik öğrenmede özgüven değişkeni tarafından açıklandığını belirtir. Yüzde açıklanan varyans değerinin (0.23) karekökü alındığında ($\sqrt{0.23}$) elde edilen 0.48 değeri, matematik öğrenmede özgüven değişkeni ile matematik

başarısı değişkeni arasındaki korelasyon katsayısını verir. Bu değer etki büyüklüğü açısından yorumlandığında, matematik öğrenmede özgüvenin etkisinin orta düzeyde olduğu söylenebilir.

Madde güçlük parametre tahminleri ve bu tahminlerin standart hataları Tablo 2’de verilmiştir. Madde parametre tahminleri lojit ölçeğinde yaklaşık 1.50 ile 6.49 aralığında değişmektedir. Ancak bu değerler madde güçlük parametre değerlerine eşit değildir. Madde güçlük değerlerini elde etmek için öncelikle öğrenci etkilerinin ortalama değerini hesaplamak gerekir ki bu değer de 5 değerın toplanması ile elde edilir: ortalama cinsiyet (0.51) çarpı cinsiyetin etkisi (-0.09); ortalama matematiğe karşı olumlu tutum puanı (12.88) çarpı matematiğe karşı olumlu tutumun etkisi (-0.01); ortalama matematiğe önem verme puanı (14.06) çarpı matematiğe önem vermenin etkisi (0.02); ortalama matematik öğrenmede özgüven puanı (10.71) çarpı matematik öğrenmede kendine güvenin etkisi (0.20) ve hata teriminin ortalaması (0). Bu değerler toplandığında elde edilen 2.24 değeri madde parametre tahminlerinden çıkarıldığında madde güçlük parametreleri elde edilir. Madde güçlük parametreleri lojit ölçeğinde yaklaşık -1.29 ile 3.69 değerleri arasında değişmekte olup ortalama madde güçlüğü yaklaşık 0.60’tır (Tablo 2). Bulunan bu değerler, Rasch modeli ile elde edilen değerlere oldukça yakındır.

Doğrusal Lojistik Test Modeli

Üç kategorili bir değişken olan bilişsel alanın, akıl yürütme kategorisi kontrol grup alınarak oluşturulan bilgi – akıl yürütme (BİLGİ) ile uygulama – akıl yürütme (UYGULA) ve dört kategorili bir değişken olan konu alanın geometri kategorisi kontrol grup alınarak oluşturulan sayılar – geometri (SAYI), cebir – geometri (CEBİR) ile veri analizi ve olasılık – geometri (VERİ)) değişkenlerinin madde güçlüğü üzerindeki etkilerinin tahminleri ile bu tahminlerin standart hataları, *t*-değerleri ve *p*-değerleri Tablo 4’te verilmiştir.

Tablo 4.

Doğrusal Lojistik Test Modeli İçin Madde Parametre Tahminleri

Madde Özellikleri	Parametre	Tahmin	Standart Hata	<i>df</i>	<i>t</i> -değeri (<i>p</i> -değeri)
BİLGİ	β_1	1.01	0.07	318	14.43 (<0.001)
UYGULA	β_2	0.33	0.07	318	4.71 (<0.001)
SAYI	β_3	-0.14	0.08	318	-1.75 (0.08)
CEBİR	β_4	0.39	0.06	318	6.50 (<0.001)
VERİ	β_5	0.30	0.07	318	4.29 (<0.001)
SABİT	β_0	-1.19	0.09	318	-13.22 (<0.001)

Bilişsel alan değişkeni için ilk kontrastın etkisi lojit ölçeğinde 0.07 standart hata ile yaklaşık 1.01 olarak tahmin edilmiştir. Bu etki 0.05 alfa düzeyinde istatistiksel olarak anlamlıdır ($p < 0.001$). Testteki akıl yürütme gerektiren maddeler, bilgi gerektiren maddelerden ortalama olarak yaklaşık 1.01 lojit daha zordur. Lojit ölçeğindeki 1.01 değerinin üsseli alındığında ($\exp(1.01)$) elde edilen 2.75 değeri odds oranını verir. O halde, bilgi gerektiren maddeler ile karşılaştırıldığında, akıl yürütme gerektiren maddeler için odds oranı yaklaşık 2.75 kat düşecektir. Diğer bir ifade ile bilgi gerektiren maddeler için maddeyi doğru olarak cevaplama olasılığı 0.50 ise, akıl yürütme gerektiren maddeler için bu olasılık yaklaşık 0.27’ye düşecektir. İkinci kontrastın etkisi lojit ölçeğinde 0.07 standart hata ile yaklaşık 0.33 olarak tahmin edilmiştir. Bu etki 0.05 alfa düzeyinde istatistiksel olarak anlamlıdır ($p < 0.001$). Testteki akıl yürütme gerektiren maddeler, uygulama gerektiren maddelerden ortalama olarak yaklaşık 0.33 lojit daha zordur. Lojit ölçeğindeki 0.33 değerinin üsseli alındığında ($\exp(0.33)$) elde edilen 1.39 değeri odds oranını verir. O halde, uygulama gerektiren maddeler ile karşılaştırıldığında, akıl yürütme gerektiren maddeler için odds oranı yaklaşık 1.39 kat düşecektir. Diğer bir ifade ile uygulama gerektiren maddeler için maddeyi doğru olarak cevaplama olasılığı 0.50 ise, akıl yürütme gerektiren maddeler için bu olasılık yaklaşık 0.42’ye düşecektir. Modelde bulunmayan bilgi – uygulama kontrastının tahmini etkisini ve standart hatasını elde etmek için ya bilişsel alan değişkeninin bilgi veya uygulama

kategorilerinden biri referans alınarak kodlama yeniden yapıp analiz tekrarlanabilir ya da ilgili değerler elde hesaplanabilir (Cohen, Cohen, West ve Aiken, 2003).

Konu alanı değişkeni için ilk kontrastın etkisi lojit ölçeğinde 0.08 standart hata ile yaklaşık -0.14 olarak tahmin edilmiştir. Bu etki 0.05 alfa düzeyinde istatistiksel olarak anlamlı değildir ($p = 0.094$). Testteki sayılar ile ilgili maddelerin güçlüğü ile geometri ile ilgili maddelerin güçlüğü arasında istatistiksel olarak anlamlı bir fark bulunmamaktadır. İkinci kontrastın etkisi lojit ölçeğinde 0.06 standart hata ile yaklaşık 0.39 olarak tahmin edilmiştir. Bu etki 0.05 alfa düzeyinde istatistiksel olarak anlamlıdır ($p < 0.001$). Testteki geometri ile ilgili maddeler, cebir ile ilgili maddelerden ortalama olarak yaklaşık 0.39 lojit daha zordur. Lojit ölçeğindeki 0.39 değerinin üsseli alındığında ($\exp(0.39)$) elde edilen 1.48 değeri odds oranını verir. O halde, cebir ile ilgili maddeler ile karşılaştırıldığında, geometri ile ilgili maddeler için odds oranı yaklaşık 1.48 kat düşecektir. Diğer bir ifade ile cebir ile ilgili maddeler için maddeyi doğru olarak cevaplama olasılığı 0.50 ise, geometri ile ilgili maddeler için bu olasılık yaklaşık 0.40'a düşecektir. Üçüncü kontrastın etkisi lojit ölçeğinde 0.07 standart hata ile yaklaşık 0.30 olarak tahmin edilmiştir. Bu etki 0.05 alfa düzeyinde istatistiksel olarak anlamlıdır ($p < 0.001$). Testteki geometri ile ilgili maddeler, veri analizi ile ilgili maddelerden ortalama olarak yaklaşık 0.30 lojit daha zordur. Lojit ölçeğindeki 0.30 değerinin üsseli alındığında ($\exp(0.39)$) elde edilen 1.35 değeri odds oranını verir. O halde, veri analizi ile ilgili maddeler ile karşılaştırıldığında, geometri ile ilgili maddeler için odds oranı yaklaşık 1.35 kat düşecektir. Diğer bir ifade ile veri analizi ile ilgili maddeler için maddeyi doğru olarak cevaplama olasılığı 0.50 ise, geometri ile ilgili maddeler için bu olasılık yaklaşık 0.43'e düşecektir.

Modeldeki sabit katsayının değeri ise 0.09 standart hata ile yaklaşık -1.19 olarak bulunmuştur. Bu değer -1 ile çarpıldığında elde edilen 1.19 değeri geometri ile ilgili akıl yürütme gerektiren maddelerin ortalama madde güçlüğünü verir.

Bu modele göre her bir maddenin tahmini güçlüğünü hesaplamak için her maddeye karşılık gelen iki madde özelliğinin modelde tahmin edilen etkisinin değerini ve sabit katsayının değerini toplamak ve toplama sonucu bulunan değeri -1 ile çarpmak gerekir. Örneğin, birinci madde uygulama gerektiren cebir ile ilgili bir maddedir. Bu durumda, bu maddenin tahmini güçlüğü 0.33 (UYGULA) + 0.39 (CEBİR) - $1.19 = -0.47 \times -1 = 0.47$ değerine eşittir. Rasch modeline göre bu maddenin güçlüğü -0.43 olarak tahmin edilmiştir (Tablo 2). Birinci maddenin doğrusal lojistik test modeline ve Rasch modeline göre tahmin edilen güçlüğü birbirine yakın değerler değildir. Bunun nedeni, doğrusal lojistik test modeline eklenen bilişsel alan ve konu alanı değişkenlerinin madde güçlüğünü açıklamak için yeterli olmamasıdır.

Bu modelde geriye kalan varyans lojit ölçeğinde yaklaşık 1.21 olarak tahmin edilmiş olup bu tahminin standart hatası 0.12'dir ve 0.05 alfa düzeyinde istatistiksel olarak anlamlıdır (t -değeri = $1.21/0.12 = 10.08$, $t_{(0.05, 318)} = 1.96$; $p < 0.001$). Bu varyans değeri, Rasch modelinde elde edilen değerden (1.71) düşüktür. Bu durum, öğrencilerin matematik başarılarındaki farklılığın madde özelliklerinden de etkilendiğini ortaya koymaktadır.

Örtük Regresyon Doğrusal Lojistik Test Modeli

Birey özelliklerinden cinsiyet (CİNSİYET), matematiğe karşı olumlu tutum (TUTUM), matematiğe önem verme (ÖNEM) ve matematik öğrenmede özgüven (ÖZGÜVEN), öğrencilerin matematik başarıları üzerindeki etkilerinin tahminleri ile bu tahminlerin standart hataları, t -değerleri ve p -değerleri Tablo 5'te verilmiştir. Birey özelliklerinin etkileri örtük regresyon Rasch modelindeki etkilere oldukça yakındır.

Madde özelliklerinden bilişsel alan ile konu alanının madde güçlüğü üzerindeki etkilerinin tahminleri ile bu tahminlerin standart hataları, t -değerleri ve p -değerleri Tablo 6'da verilmiştir. Madde özelliklerinin etkileri doğrusal lojistik test modelindeki etkilere çok yakındır.

Tablo 5.

Örtük Regresyon Doğrusal Lojistik Test Modeli İçin Öğrenci Parametre Tahminleri

Öğrenci Özellikleri	Parametre	Tahmin	Standart Hata	df	t-değeri (p-değeri)
CİNSİYET	ψ_1	-0.08	0.12	318	-0.67 (0.50)
TUTUM	ψ_2	-0.01	0.03	318	-0.33 (0.74)
ÖNEM	ψ_3	0.02	0.03	318	0.67 (0.50)
ÖZGÜVEN	ψ_4	0.17	0.02	318	8.50 (<0.001)

Tablo 6.

Örtük Regresyon Doğrusal Lojistik Test Modeli İçin Madde Parametre Tahminleri

Madde Özellikleri	Parametre	Tahmin	Standart Hata	df	t-değeri (p-değeri)
BİLGİ	β_1	1.01	0.07	318	14.43 (<0.001)
UYGULA	β_2	0.33	0.07	318	4.71 (<0.001)
SAYI	β_3	-0.14	0.08	318	-1.75 (0.08)
CEBİR	β_4	0.39	0.06	318	6.50 (<0.001)
VERİ	β_5	0.30	0.07	318	4.29 (<0.001)
SABİT	β_0	-3.09	0.45	318	-6.87 (<0.001)

Cinsiyet, matematiğe karşı olumlu tutum, matematiğe önem verme ve matematik öğrenmede özgüvenin öğrencilerin matematik başarıları üzerindeki etkileri hesaba alındıktan sonra geriye kalan birey varyansı 0.09 standart hata ile 0.93 olarak tahmin edilmiştir.

Tartışma

Geleneksel madde tepki kuramı (MTK) modelleri yalnızca madde ve birey parametrelerini tahmin etmekle sınırlıdır. Bu modeller, maddeler ve/veya bireyler arasındaki farklılıkları açıklamak üzere modele madde ve/veya birey özelliklerinin eklenmesine izin verecek kadar esnek değillerdir. Diğer yandan açıklayıcı madde tepki modellerinin, geleneksel MTK modellerine göre böyle bir esnekliği vardır. Bu çalışmada, açıklayıcı madde tepki modellerinin bahsedilen avantajı birey özellikleri ile sekizinci sınıf Türk öğrencilerinin matematik başarıları arasında gözlenen farklılıkları açıklamak ve madde güçlüklerinin madde özelliklerine göre nasıl değiştiğini incelemek için kullanılmıştır. Ayrıca birey özellikleri ile madde özellikleri çarpılarak elde edilen birey-madde yordayıcıları, farklılaşan madde işlevi modellerini tanımlamada kullanılabilirler.

Bu çalışmada sadece iki kategorili puanlanan maddelerden oluşan bir test kullanılmış ve iki kategorili madde tepki kuramı modellerinden Rasch modeli ele alınmıştır. Bu çalışmada kullanılan modelleri iki-parametrelili veya üç-parametrelili lojistik modellere uygulamak mümkündür. Ayrıca bağlantı fonksiyonu olarak lojit bağlantısı yerine probit bağlantısı kullanıldığında normal ojiv modelleri de elde edilebilir. Çok kategorili puanlanan maddelerden oluşan testler veya ölçekler için de kısmi puanlama modeli, aşamalı tepki modeli vb. kullanılabilir.

Açıklayıcı madde tepki modellerinin analizi için kullanılacak çeşitli program veya yazılımlar bulunmaktadır. Bu çalışmada STATA programı üzerinden çalışan GLLAMM yazılımı kullanılmıştır. SAS NLMIXED, SAS GLIMMIX, HLM, MLwiN gibi programlar da açıklayıcı madde tepki modellerinin analizinde kullanılabilir programlar arasındadırlar (Açıklayıcı madde tepki kuramı modellerinin analizinde kullanılan programlar hakkında daha ayrıntılı bilgi edinmek için bakınız De Boeck ve Wilson, 2004).

Sonuçlar

TIMSS 2007 Türkiye sekizinci sınıf matematik verisi ilk olarak açıklayıcı madde tepki modelleri altında Rasch modeli ile analiz edilmiştir. Birey kısmında, öğrencilerin matematik başarılarının varyansı yaklaşık 1.71 (0.16) lojit olarak tahmin edilmiş olup matematik başarıları arasında gözlenen farklılıklar 0.05 alfa düzeyinde istatistiksel olarak anlamlı bulunmuştur. Madde kısmında ise, madde güçlüklerinin yaklaşık -1.31 ile 3.67 lojit arasında değiştiği, ortalama madde güçlüğü yaklaşık 0.58 lojit olduğu bulunmuştur. Bu ortalama madde güçlüğü değeri, öğrencilerin ortalama matematik başarıları değeri olan sıfırdan yüksek olup, ortalama matematik başarılarına sahip bir öğrencinin teste orta güçlükteki bir maddeyi doğru olarak cevaplama olasılığı %50'den düşüktür. O halde, testin zor bir test olduğu söylenebilir.

Rasch modeli ile öğrencilerin matematik başarıları arasında anlamlı farklılıklar bulunduktan sonra, öğrenciler arasındaki farklılıkları açıklamak üzere modelin birey kısmına öğrenci özelliklerinden cinsiyet, matematiğe karşı olumlu tutum, matematiğe önem verme ve matematik öğrenmede özgüven değişkenleri yordayıcı değişkenler olarak eklenerek birey kısmında bir regresyon modeli oluşturulmuş, modelin madde kısmı ise aynı kalmıştır. Örtük regresyon Rasch modeli olarak adlandırılan bu modelin analizi sonucunda, öğrencilerin matematik başarılarında gözlenen farklılıkların cinsiyetten, matematiğe önem verip, matematiğe karşı olumlu bir tutuma sahip olmaktan etkilenmediği, ancak bu farklılıkların %23'lük bir kısmının öğrencilerin matematik konusunda kendilerine olan güvenleri ile açıklandığı bulunmuştur. Öğrencilerin matematik konusunda kendilerine olan güvenleri arttıkça, matematik başarıları da istatistiksel olarak anlamlı bir şekilde artmaktadır. Öğrencilerin matematik başarıları arasındaki farklılıkların bir kısmı modele eklenen öğrenci özellikleri ile açıklandıktan sonra geriye kalan varyans değeri yaklaşık 1.32 (0.13) lojit olarak tahmin edilmiş olup model ile açıklanmayan bu varyans değeri 0.05 alfa düzeyinde istatistiksel olarak anlamlıdır. Bu model ile elde edilen madde güçlük değerleri, Rasch modeli ile elde edilen değerlere çok yakındır.

Rasch modelinin birey kısmı aynı kalıp modelin madde kısmına madde güçlükleri arasındaki farklılıkları açıklamak üzere madde özelliklerinden bilişsel alan ve konu alanı değişkenleri yordayıcı değişkenler olarak eklenerek doğrusal lojistik test modeli oluşturulmuştur. Bu modelin analizi sonucunda, bilişsel alan ile konu alanının madde güçlüğü üzerinde etkili olduğu bulunmuştur. Bilişsel alan değişkeni incelendiğinde, akıl yürütme gerektiren maddelerin bilgi gerektiren veya uygulama gerektiren maddelerden daha zor olduğu, bilgi veya uygulama gerektiren maddeler ile akıl yürütme gerektiren maddeler arasında madde güçlüğü bakımından istatistiksel olarak anlamlı bir fark olduğu bulunmuştur. Konu alanı değişkeni incelendiğinde, geometri ile ilgili maddelerin cebir veya veri analizi ve olasılık ile ilgili maddelere göre daha zor olduğu, geometri ile ilgili maddeler ve sayılar ile ilgili maddeler arasında ise madde güçlüğü bakımından istatistiksel olarak anlamlı bir fark olmadığı bulunmuştur.

Örtük regresyon doğrusal lojistik test modeli ile bulunan sonuçlar, örtük regresyon Rasch modeli ve doğrusal lojistik test modeli ile bulunan sonuçlara çok yakındır. Hem bireyler hem de madde güçlükleri arasında gözlenen farklılıkların açıklanmaya çalışılması durumunda, örtük regresyon Rasch modeli ve doğrusal lojistik test modeli analizleri ile iki ayrı analiz yapmak yerine, örtük regresyon doğrusal lojistik test modeli analizi ile tek bir analiz yapılabilir.

Kaynakça

- Andrich, D. (1978a). Application of a psychometric model to ordered categories which are scored with successive integers. *Applied Psychological Measurement*, 2, 581-594.
- Andrich, D. (1978b). A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.
- Breslow, N. E. & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88, 9-25.

- Briggs, D. C. (2008). Using explanatory item response models to analyze group differences in science achievement. *Applied Measurement in Education*, 21, 89-118.
- Cohen, J., Cohen, P., West, S. G. & Aiken L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum.
- Davidian, M. & Giltinan, D. M. (1995). *Nonlinear models for repeated measurement data*. London: Chapman & Hall.
- De Boeck, P. & Wilson, M. (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. New York: Springer
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.
- IEA (2005). *TIMSS 2007 Assessment Framework*. Boston College, MA:IEA
- Lord, F. N. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Lord, F. N. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.
- McCulloch, C. E. & Searle, S. R. (2001). *Generalized, linear, and mixed models*. New York: Wiley.
- Rabe-Hesketh, S., Skrondal, A. & Pickles, A. (2004). *GLLAMM manual*. U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 160.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph*, No. 17.
- Schwarz, G. (1978). Estimating a dimension of a model. *Annals of Statistics*, 6, 461- 464.
- Skrondal, A. & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal and structural equation models*. Boca Raton: Chapman & Hall/CRC.
- StataCorp. (2003). *Stata statistical software: Release 8.0*. College Station, TX: Stata Corporation.