Turkish Studies
Volume 14 Issue 1, 2019, p. 307-332
DOI: 10.7827/TurkishStudies.14745
ISSN: 1308-2140
Skopje/MACEDONIA-Ankara/TURKEY

INTERNATIONAL BALKAN UNIVERSITY
EXCELLENCE FOR THE FUTURE
IBU.EDU.MK

Research Article / Araştırma Makalesi

Article Info/Makale Bilgisi

Received/Geliş: *Aralık 2018* ***Accepted/Kabul:*** *Şubat 2019*

Referees/Hakemler: *Prof. Dr. Adnan KAN – Doç. Dr. Kaan Zülfikar DENİZ*

This article was checked by iThenticate.

BİR, İKİ, ÜÇ VE DÖRT PARAMETRELİ LOJİSTİK MADDE TEPKİ KURAMI MODELLERİNİN KARŞILAŞTIRILMASI*

*Ayşenur ERDEMİR**- Emine ÖNEN****

ÖZET

Bu araştırma kapsamında 1PLM, 2PLM, 3PLM ve 4PLM'nin araştırma verisine uyum düzeyleri incelenmiş, bu modeller altında ayrı ayrı kestirilen madde ve yetenek parametrelerinin doğruluğu karşılaştırılmış, maddelerin ve ilgili alt testin toplam olarak bu modeller altında sağladıkları bilgi miktarları hesaplanmış ve karşılaştırılmıştır. Bu karşılaştırmaların yapılmasında MEB'den temin edilen 2012 SBS Türkçe alt testi verileri kullanılmıştır. Bu verilerden seçkisiz olarak seçilen 1500 yanıtlayıcı, çalışma grubunu oluşturmaktadır. Araştırma sorularına yönelik analize geçilmeden önce MTK varsayımları tek boyutluluk ve yerel bağımsızlık test edilmiştir. Tek boyutluluk varsayımı için tetrakorik korelasyon matrisine dayalı AFA yapılmış testin tek boyutlu olduğu sonucuna ulaşılmıştır. Yerel bağımsızlık varsayımının karşılanıp karşılanmadığının incelenmesinde Yen'in Q_3 istatistiği kullanılmış, ilgili bütün madde çiftleri için bütün modeller altında yerel bağımsızlık varsayımının karşılandığı görülmüştür. Model-veri uyumu bağlamında en iyi uyuma sahip modelin 4PLM olduğu bulunmuştur. Madde parametreleri R Studio programında MML kestirim yöntemi kullanılarak kestirilmiştir. Madde parametrelerinin kestirimine ilişkin standart hata değerleri oldukça küçüktür ve bu madde parametrelerinin doğru bir şekilde kestirildiğine işaret etmektedir. Her bir yanıtlayıcı için ilgili bütün modeller altında yetenek kestirimi ML kestirim yöntemi kullanılarak yapılmıştır. Yetenek kestirimine ilişkin standart hatalar ANOVA tekniğiyle karşılaştırılmıştır. Karşılaştırma sonucu, 4PLM altında yapılan kestirimin, diğer 3 modelden daha düşük standart hataya sahip olduğunu ve yetenek parametresinin bu model altında daha doğru

*Bu çalışma Ayşenur Erdemir tarafından Gazi Üniversitesi Eğitim Bilimleri Enstitüsü'nde Dr. Öğr. Üyesi Emine Önen danışmanlığında hazırlanan "Bir, iki, üç ve dört parametreli lojistik madde tepki kuramı modellerinin karşılaştırılması" isimli yüksek lisans tezinden üretilmiştir.

** Arş. Gör., Gazi Üniversitesi, E-posta: erdemiraysenur@gmail.com

*** Dr. Öğr.Üyesi, Gazi Üniversitesi, Gazi Eğitim Fakültesi, E-posta: emineonenn78@gmail.com

kestirildiğini göstermektedir. Yine 4PLM'nin ilgili veri seti için sağladığı bilgi miktarının diğerlerinden fazla olduğu ve testin amacına uygun şekilde en fazla bilgiyi orta yetenek düzeyinde verdiği görülmüştür. Bütün modeller en fazla bilgiyi θ = -1 ile θ = 0 yetenek aralığında sağlamaktadırlar.

Anahtar Kelimeler: Madde Tepki Kuramı, 4 Parametreli Lojistik Model, Parametre Kestirim Doğruluğu, Bilgi Fonksiyonları

COMPARISON OF 1PL, 2PL, 3PL AND 4 PL ITEM RESPONSE THEORY MODELS

ABSTRACT

In this study, model-data fit studies were conducted for 1PLM, 2PLM, 3PLM and 4PLM, accuracy of item and ability parameter estimation was compared, and item and test information functions were provided. To be able to compare models on these issue, 2012 SBS (high school entrance exam) Turkish subtest were taken from Ministry of Education. The study group was comprised by 1500 examinee. Before the analysis of research problems, the assumptions of IRT were checked. First, the data were checked for unidimensionality with EFA based upon tetrachoric correlation matrix, and it appeared essentially unidimensional. Then, all pairs of items were checked for local independence using Yen's Q_3. None of the pairwise residual correlations for all 4 models were greater than .20 in absolute value, which showed that local dependence did not appear to be a problem. The item and ability parameters were provided the property of invariance. In terms of model-data fit, it was found that 4PLM was the best fitting model of all. The items were calibrated with all related models using MML in R Studio. The standard errors for most of the item parameters were reasonably small. This showed that the item parameters were estimated with good accuracy. The abilities (θ) were estimated for each individual examinee with all 4 models using maximum likelihood estimation. The standard errors of ability estimate were compared using Analysis of Variance for the degree of estimation accuracy. It was found that the estimation with 4PLM had smaller standard errors than the other 3 models, and the abilities were most precisely estimated with 4PLM. In addition, the most information was provided by 4PLM for the related test. All the models provided more information on the theta interval -1 and 0 than the other θ levels.

STRUCTURED ABSTRACT

Introduction

Item response theory (IRT) prescribes model-based measurement in which ability estimates are made based on the individuals' responses to items and items' properties. Fundamentally, IRT is based on a mathematical model about how respondents at different ability levels answer to an item (Baker, 2001, p. 6; Crocker & Algina, 1986, p. 339-340; De Ayala, 2008, p. 4-5; Embretson & Reise, 2000, p. 13; Hambleton

and others, 1991, p. 5). In this study, unidimensional dichotomous IRT models have been compared in terms of the model fit, accuracy of both item parameters and ability parameters and amount of information that a test provided under a specific IRT model. The most used unidimensional dichotomous IRT models are one-parameter logistic model (1PLM), two-parameter logistic model (2PLM), three-parameter logistic model (3PLM). Four-parameter logistic model (4PLM) was developed by Barton and Lord (1981) by adding *d* parameter which indicates the probabilities of answering incorrectly of respondents at high-level ability to an item at low-difficulty level. *d* parameter allows upper asymptote to get values lower than 1.00, which enables that the location at the theta continuum does not change a considerable extent in the case of a respondent at high-level ability incorrectly responding to an easy item. Magis (2013) has stated that when ability parameters are estimated under 4PLM, both items and test provides more information.

It has been noticed that at the studies in which the dichotomous items have been investigated in Turkey, in general, 1PLM, 2PLM and 3PLM have been used but the item information function and test information function has not been calculated (Can, 2003; Kılıç, 1999; Yapar, 2003; Yeğin, 2003). Also, in Turkey, it could not be reached any study related to 4PLM. When the worldwide literature is revised, it is seen that 4PLM has been studied in the fields of psychopatlogy and CAT applications (Loken & Rulison, 2010; Reise & Waller, 2003; Rulison & Loken, 2009; Rupp, 2003; Waller & Reise, 2010). Barton and Lord (1981) estimated the ability parameter by constraining *d* parameters in their study related to 4PLM. In the current study, based on a standard achievement test data, 1PLM, 2PLM, 3PLM and 4PLM have been compared. Especially at the high-stakes tests, incorrectly responding to an easy item of the students at high-level ability because of inattention or other factors could affect their ability estimates and in that case it would be important to investigate the accuracy level of the decisions that made based on these ability estimates. For the aims of the students' selection and placement to the high schools and universities, these types of tests are administered in Turkey. Through these tests, it is crucial to be estimated more accurately of the students' ability levels and to get more information about them. The studies (Loken & Rulison, 2010; Rulison & Loken, 2009; Waller & Reise, 2010; Yen, Ho, Laio, Chen & Kuo, 2012) conducted in this subject reveals that 4PLM provides more accurate ability estimations. Accordingly, it is considered that it is important to get ability parameters based on high-stakes tests by using 4PLM.

Based on this theoretical explanations and research findings, the current study was conducted to investigate comparatively 1PLM, 2PLM, 3PLM and 4PLM in terms of the model fit, accuracy of the model parameters' estimates, item information function and test information function based on the data obtained Turkish Subtest of Selection and Placement Examination for the high school administered in 2012. It is considered that this study is contributed to the related literature in terms of (a) freely estimation of *d* parameters (unlike constraining this parameter as Barton and Lord done in their study) and (b) using 4PLM for a national large-scale achievement test differently from the studies related to 4PLM in which this model was used for CAT application.

Method

Study Group

This study was carried out based on the data obtained Turkish Subtest of Selection and Placement Examination for the high school administered to 8th grade students in 2012. The study group was comprised of 1500 students who were randomly chosen from the data set belonging to 50000 students which were get for this study from the Ministry of National Education, Department of Measurement, Evaluation and Examination Services.

Procedure and Data Analysis

In this study, initially, the analyses were run to obtain evidences for validity and reliability of the related measures. By exploratory factor analyses (EFA) which was applied to obtain evidence in terms of construct validity, it was also tested the unidimensionality assumption of IRT. Accordingly, based on tetrachoric correlation matrix, exploratory factor analysis with maximum likelihood factor extraction method was applied. As a result of analysis, it was observed that all respondents correctly answered the 23th item in the test (item variance=0) and the magnitude of the factor loading of 20th item was low. Therefore, EFA was rerun by excluding these items and fit-value was calculated as 0.98 and RMSEA=0.064. The eigenvalue of the first factor was found as 6.82 and the explained variance of this factor was found as 32%. These findings were considered as the evidence of validity and they also indicated that the unidimensionality assumption was met. As evidence for the reliability of the related measures, the marginal reliability coefficients were calculated as 0.820, 0.827, 0.826 and 0.831 under 1PLM, 2PLM, 3PLM and 4PLM respectively. Afterwards, in order to test local independency assumption of IRT, the Q_3 statistics for the 210 item-pairs related to 21 item in the Turkish subtest were calculated separately under each model. All the Q_3 statistics calculated for the all possible item pairs were found to be lower than 0.20 and these findings indicated that the local independency assumption was met for the related data set.

After checking the assumptions, the model fit of the four IRT models at this study were tested and compared. In order to assess model fit, –2LL, AIC, BIC and SMRSR values were considered and also based on –2LL values comparisons were made by Analysis of Variance (ANOVA). For the item fit, Bock's (1972) x^2 index was used. The item and ability parameters were estimated under each model separately. In order to estimate model parameters and standard errors Marginal Maximum Likelihood (MML) was used. In the comparison of accuracy of the parameter estimates, in order to test if the standard errors belonging to these related measures differed significantly from each other, repeated measures ANOVA was used (Field, 2005, s. 458). For the significant differences, the effect size index (η^2) was calculated and these indexes were considered according to the Cohen's (1988) criteria. By calculating test information function, it was found that under which model this test provides most information.

Results

In this study, the overall fit of each model to the related data set were tested and the fit levels of the models were compared in pairs.

Analysis results revealed that the more complex model (models with more parameters) demonstrated better fit to the related data. As a result of analysis, it was found that the best model that explained response behaviors of the students included in the study group to the items in the Turkish Subtest of Selection and Placement Examination for the high school administered in 2012 was 4PLM. When examined item fit, it was observed that as the number of estimated parameters increased, the items which did not fit the model decreased. Because it could not be found any code that test item fit for 4PLM at R statistical computing environment and any other software, it could not be tested item fit under 4PLM.

Considering the accuracy of the estimation of the item parameters, it was seen that the standard errors belonging to the item difficulty parameter estimates were so low for all IRT models tested in this study. However, the standard errors belonging to the item discrimination parameter estimates increased as the model being more complex. It was observed that the means of the standard errors belonging to the pseudo-guessing parameter estimates were found to be equal under 3PLM and 4PLM. Finally, the mean of the standard errors belonging to the *d* parameter estimates also was found as so low.

As a result of the analysis made to compare the accuracy of the ability parameter, significant differences were found between all possible model comparisons. When the results of ANOVA and calculated effect size indexes assessed together, it was considered that ability parameter was estimated with minimum error under 4PLM. Test information functions were calculated for θ=-4 and θ=4 theta interval (as a whole) and also for eight sub intervals (-4- (- 3.10); -3.02- (-2.12); -2.04- (-1.14); -1.06- (-0.08); 0.08- 1.06; 1.14- 2.04; 2.12- 3.02; 3.10- 4.00). Also the ability ranges in which all of these models provided the most information was seen to be between θ= -1,06 and θ= -0,08. As moving to the end of the theta continuum the test provided less information. It is a striking finding that the model which provided most information is 4PLM for theta interval of θ= -1,06 and θ= 1.06.

Discussion and Conclusion

In this study, unidimensional dichotomous IRT models were compared in terms of model-fit, accuracy of the model parameters, item information function and test information function. Based on the comparisons about model fit, among the tested models the relatively better fitted model was found to be 4PLM as consistent with the results of Loken ve Rulison's (2010) study. Findings related to model fit revealed that in examining the items' properties it had to be taken in to account *d* parameter which indicated the probabilities of answering incorrectly of respondents at high-level ability to an item at low-difficulty level because of inattention or other factors.

When looking at the accuracy of the item parameters it was seen that the standard errors belonging to the item difficulty parameters were low, the standard errors belonging to the item discrimination parameters increased as the models became more complex, the accuracy levels of the *c* parameters under 3PLM and 4PLM were high and similar, and *d* parameters were estimated with a small amount of error under 4PLM. It is considered that the increase in the standard error values belonging to

the item discrimination parameters could be related to the constraints made on *a* parameters in estimation process. Item parameters enables information about item's quality. In this study, estimates of item difficulty parameter under all four IRT models indicate that most of the items in the test are easy contrary to what is expected to be in a test to determine students' levels. When the item discrimination parameters were examined it was seen that the highest estimates were obtained under 4PLM. It was observed that the *c* parameters for many items under 3PLM and 4PLM were nonzero values. It is expected that the *c* parameter estimations are low when the distractors of the test items are strong and they are chosen by respondents at low ability level (Lord, 1974). In this context, it could be concluded that the distractors of the items in the current dataset do not function as expected and there is a possibility to answer these items by chance. For many items *d* parameters were observed to get values lower than 1.00. These findings indicate that respondents with high ability level go the options other than correct answer because of the inattention or other factors. This situation may be due to complex expressions or an incoherent statement in the item stem.

In this study it was observed that the ability parameter was estimated with smallest amount of error under 4PLM. The fact that the highest value of the item discrimination parameters for the items are obtained under 4PLM and also getting the most information under this model support this finding. In consistent with current study results, in some other studies (Liao, Ho, & Yen, 2012; Rulison & Loken, 2009; Yen, Ho, Liao, & Chen, 2012; Yen, Ho, Liao, Chen, & Kuo, 2012) it was found that when the 4PLM was used instead of 3PLM the standard error of estimation decreased and therefore the accuracy of the estimation increased. Test information functions calculated under each of the four IRT models indicated that this test provided the most information for slightly below the medium ability range. Because Selection and Placement Examination for the High School is administered to determine students' achievement levels, it is expected that there are more items at medium difficulty level in this test and also in this context this test is supposed to provide more information at medium ability level. In line with this expectation, in the current study it was found that the test provided the most information at medium ability level under 4PLM. When the test information functions and standard errors of parameter estimates are assessed together, it is concluded that among four IRT models, 4PLM provides the most information and the model parameters are estimated with the small amount of error. Similarly, Waller and Reise (2010) stated that the most information about respondents' ability was provided under 4PLM. Accordingly, it is understood that it is important and necessary to use 4PLM that takes into account *d* parameter especially for the situations in which inferences are made about individuals' ability and decisions are taken about them.

Keywords: Item Response Theory, 4 Parameter Logistic Model, Accuracy of Estimation, Information Functions

1. Giriş

Sosyal bilimlerden biri olan eğitim çalışmaları temelde insan davranışları ile ilgilenmektedir. İnsan davranışlarını açıklamaya yönelik olarak da psikolojik yapılar incelenmektedir. Eğitimin çalışma

alanları da sosyal bilimlerin diğer alanlarında olduğu gibi psikolojik yapıların açıklanması, analiz edilmesi ve yorumlanmasını gerektirmektedir. Fakat psikolojik yapılar dolaylı yoldan ölçülebilmekte ve bu da bazı problemleri beraberinde getirmektedir (Baykul, 2010, s. 95; Turgut ve Baykul, 1992, s. 3-4). Test kuramı, bu problemlerin ve çözümlerinin çalışılmasına dayanmaktadır.

Gerçek puan modeli olarak da anılan Klasik Test Kuramı (KTK), gözlenen puanın, gerçek puan ve seçkisiz hata bileşenlerinden oluştuğunu varsaymaktadır (Spearman 1904, 1907). Bir diğer kuram, Örtük Özellik Kuramı olarak da bilinen Madde Tepki Kuramı (MTK), hem kişinin yanıtlarına hem de uygulanan maddelerin niteliklerine dayanarak yetenek kestirimi yapan model temelli bir ölçmedir (Lord ve Novick'in 1968, s. 358). MTK modelleri, maddeye verilen yanıtların hepsinin altında yatan bir örtük özellik (theta: θ) olduğunu varsaymaktadır. Bu örtük özellik ve madde nitelikleri arasındaki ilişki tanımlanmakta ve bir ölçek üzerinde bir maddenin yanıtlanma olasılığı kestirilmektedir. Madde Tepki Kuramının öne çıkan bir takım özellikleri, KTK'den farkını önemli ölçüde ortaya koymaktadır: (a) Madde nitelikleri gruba bağımlı değildir. (b) Yetenek puanları teste bağımlı değildir. (c) Model test düzeyinde değil madde düzeyinde açıklanır. (d) Güvenirliği kestirmek için paralel testlere ihtiyaç duyulmaz. (e) Her bir yetenek puanı için bir kesinlik ölçüsü (a measure of precision) sağlar (Baker, 2001, s. 6; Crocker ve Algina, 1986, s. 339-340; De Ayala, 2008, s. 4-5; Embretson ve Reise, 2000, s. 13; Hambleton vd., 1991, s. 5).

MTK'nin temel varsayımları; tek boyutluluk, yerel bağımsızlıktır. Bunlardan ilki test ile ölçülen özelliğin tek boyutlu olduğu varsayımıdır (Embreson ve Reise, 2000, s. 48). Yerel bağımsızlık ise aynı θ (θ_1, θ_2, … , θ_k) değerlerine sahip bireylerin oluşturduğu bir yanıtlayıcı grubu için maddelere doğru yanıt verme olasılık değerlerinin birbirinden bağımsız oluşu olarak tanımlanmaktadır (Lord ve Novick 1968, s. 360-362).

Madde Tepki Kuramı ile ilgili temel kavramlar olan madde bilgi fonksiyonu(MBF) ve test bilgi fonksiyonu(TBF), maddeleri ve testleri tanımlamada, maddeleri seçmede ve testleri karşılaştırmada MTK'nin güçlü olmasını sağlayan unsurlardandır. Madde bilgi fonksiyonu her bir yetenek düzeyinde ilgili maddenin bireyin yeteneğinin kestirimi konusunda ne kadar bilgi sağladığını göstermektedir. MBF, $I_i(\theta)$, matematiksel olarak aşağıdaki eşitlik ile gösterilir:

$$I_i(\theta) = \frac{[\acute{P}_i(\theta)]^2}{P_i(\theta)\, Q_i(\theta)} \qquad i = 1, 2, \ldots, n \qquad \text{(Eşitlik 1)}$$

Burada, $I_i(\theta)$, *i.* maddenin madde bilgi fonksiyonunu; $P_i(\theta)$, θ yetenek düzeyindeki bir bireyin *i.* maddeye tepki verme olasılığını (madde tepki fonksiyonu-MTF); $Q_i(\theta)$, $1 - P_i(\theta)$'yı ve $\acute{P}_i(\theta)$, $P_i(\theta)$'nın birinci türevini göstermektedir θ yetenek düzeyinde test bilgi fonksiyonu ($I\theta$) ise, θ düzeyindeki madde bilgi fonksiyonlarının toplamıdır ve aşağıdaki eşitlik ile gösterilir:

$$I(\theta) = \sum_{i=1}^{n} I_i(\theta) \qquad \text{(Eşitlik 2)}$$

Maddeler bağımsız olarak test bilgi fonksiyonuna katkı sağlarlar. Her bir test maddesinin katkısı, testteki diğer maddelerin katkısından bağımsız olarak belirlenebilir. Bu durum KTK kapsamında mümkün değildir; test maddelerinin testin güvenirliğine ve madde ayırt edicilik indekslerine katkısı, testteki diğer maddelerin özelliklerinden bağımsız bir şekilde belirlenemez.

Bir maddenin uygulandığı bireyin yetenek düzeyi θ, maddenin güçlük parametresine eşit olduğunda yeteneğin en doğru şekilde ölçüldüğü ve sağladığı bilgi miktarının fazla olduğu şeklinde yorumlanmaktadır. Yani, θ ölçeği üzerinde yetenek parametresi madde güçlük düzeyinden uzaklaştıkça maddenin sağladığı bilginin miktarı azalmakta ve yetenek ölçeğinin uç noktalarında sıfıra yaklaşmaktadır. Test bilgi fonksiyonunun ölçek üzerindeki düzeyi tek bir maddenin bilgi fonksiyonunun ölçek üzerindeki düzeyinden çok daha yüksek olmaktadır. Temel olarak test bilgi fonksiyonu, yetenek ölçeğinin tüm ranjını dikkate alarak testin yetenek kestiriminde ne kadar iyi olduğu bilgisini vermektedir. İstenilen test bilgi fonksiyonuna ait eğrinin yatay olduğu ve bütün yetenek düzeyleri için

birbirine yakın oranlarda bilgi sağlaması gerektiği ifade edilirken, bunun tüm durumlar için en iyisi olduğu söylenememektedir. Özetle, test bilgi fonksiyonu testin amacına göre değerlendirilmelidir. Örneğin ödüllü bir sınavda kesme puanının olduğu noktada zirve yapan test bilgi fonksiyonunun en iyi olduğu ifade edilir (Baker, 2001, s. 108-109; Hambleton vd., 1991, s. 91-95).

Test bilgi fonksiyonu ile ilişkili bir diğer kavram hatadır. TBF, testte yer alan bütün maddelerin o maddeler ile ölçülen örtük özelliğe ilişkin olarak ne düzeyde bilgi verdiğini gösterir. Bu fonksiyona dayalı olarak yetenek parametresi kestiriminin ne kadar az hata içerdiği de belirlenebilmektedir (Hambleton ve Swaminathan, 1985, s. 104-105). Yetenek parametresi kestiriminin standart hatası Eşitlik 3'te verilmiştir:

$$SE(\hat{\theta}) = \frac{1}{\sqrt{I(\theta)}} \qquad \text{(Eşitlik 3)}$$

MTK bağlamında $SE(\hat{\theta})$, klasik ölçme kuramındaki ölçmenin standart hatası ile aynı amaca hizmet eder. Fakat $SE(\hat{\theta})$ değeri yetenek düzeyine göre değişirken, KTK bağlamında hesaplanan ölçmenin standart hatası değişmemektedir.

1.1. Model veri uyumu

MTK çerçevesinde modellemeden faydalanabilmek için seçilen MTK modellerinin veriye yeterli düzeyde uyum göstermesi gerekmektedir. Dolayısıyla model-veri uyumunun değerlendirilmesi, model geçerliğinin bir parçasıdır. Bir MTK modeli ile veri arasındaki uyumsuzluk; model varsayımlarının karşılanamaması, madde ve yetenek kestirimlerindeki değişmezlik özelliğinin elde edilememesi, uygun MTK modelinin seçilmemesi ya da örneklem büyüklüğünün çok küçük olması gibi nedenlerden kaynaklanabilmektedir (Hambleton ve Jones, 1993).

1.2. Madde tepki kuramında modeller

1.2.1. 1 parametreli lojistik model (1plm)

1 parametreli lojistik model, en temel modeldir. Bu modelde yanıtlayıcının bir maddeye doğru yanıt verme olasılığı, yanıtlayıcının yetenek düzeyine (θ) ve o maddenin güçlük düzeyine (*b* parametresi) bağlı olarak hesaplanmaktadır. 1PLM'de, şans başarısının ve dikkatsizlik durumunun olmadığı, ayırt edicilik parametresinin (*a*) ise maddeler arasında farklılaşmadığı varsayılmaktadır. 1PLM için bir yanıtlayıcının i maddesine doğru yanıt verme olasılığı Eşitlik 4'te sunulmaktadır:

$$P_i(\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}} \qquad \text{(Eşitlik 4)}$$

$P_i(\theta)$, θ yetenek seviyesinde seçkisiz olarak seçilen bir yanıtlayıcının *i* maddesine doğru olarak yanıt verme olasılığıdır. b_i, *i* maddesinin güçlük parametresi ve *e* değeri 2,718 olan sabit bir sayıdır. b_i parametresinin sayısal değeri ne kadar büyükse, maddeyi doğru yanıtlamak için bireyin sahip olması gereken yetenek düzeyinin o kadar yüksek olması gerektirmektedir. Bu da maddenin daha zor bir madde olduğuna işaret etmektedir. Zor maddeler yetenek ölçeğinin sağında ya da yüksek ucunda gösterilirken; kolay maddeler ise solda ya da alt uçta gösterilmektedir. b_i parametresi genellikle -2,00 ile +2,00 arasında değerler alır; b_i değeri -2,00'a yakın maddeler çok kolay, +2,00'a yakın maddeler çok zordur (Hambleton vd., 1991, s. 13; Hambleton ve Swaminathan, 1985, s. 36).

1.2.2. 2 parametreli lojistik model (2plm)

Birnbaum (1968) 2 parametreli lojistik modeli geliştirmiştir. 2PLM'de, 1PLM'nin bir adım ötesine giderek madde güçlük parametresine ek olarak madde ayırt edicilik indeksi (*a* parametresi) de analize dâhil edilmiştir. Fakat alt ve üst asimptotların (*c* – *d* parametreleri) sıfır olduğu varsayılmaktadır. 2PLM için bir yanıtlayıcının *i* maddesine doğru yanıt verme olasılığı Eşitlik 5'te sunulmaktadır:

$$P_i(\theta) = \frac{e^{D_{ai}(\theta - b_i)}}{1 + e^{D_{ai}(\theta - b_i)}} \qquad \text{(Eşitlik 5)}$$

1PLM'den farklı olarak iki yeni öğe vardır. Bunlar; lojistik fonksiyonu normal ogive fonksiyonuna yakınlaştırmak için eklenen bir ölçekleme faktörü olan *D* faktörü ve madde ayırt edicilik parametresi olan a_i'dir. Ölçekleme faktörü genel olarak $D = 1{,}7$ olarak kabul edilir. Teorik olarak madde ayırt edicilik parametresi değeri -∞, +∞ arasında değişse de bu parametrenin uygulamalarda sıklıkla a_i = 0,00 ile a_i = 2,00 arasında değerler almaktadır. a_i parametre değeri 0'a yaklaşan maddelerin farklı yetenek düzeyindeki bireyleri ayırt etme gücü düşüktür. a_i'nin sayısal değeri yükseldikçe ayırt edicilik de yükselmektedir (Hambleton vd., 1991, s. 15).

1.2.3. 3 parametreli lojistik model (3plm)

Birnbaum (1968) 2PLM'ye, doğru yanıt verme olasılığına şans başarısının katkısını gösteren bir parametre ekleyerek 3 parametreli lojistik modeli geliştirmiştir. 3PLM, ilk iki modele ek olarak c_i (pseudo-chance-level) parametresini içermektedir. 1- ve 2-parametreli lojistik modellerde alt asimptot sıfır olarak kabul edilmekte ve şansla doğru yanıt verme olasılığı dikkate alınmamaktadır. 3PLM ise, madde karakteristik eğrisinde sıfırdan farklı alt asimptot imkânı sunmaktadır. 3PLM için bir yanıtlayıcının *i* maddesine doğru yanıt verme olasılığı Eşitlik 6'da sunulmaktadır:

$$P_i(\theta) = c_i + (1 - c_i)\frac{e^{D_{ai}(\theta - b_i)}}{1 + e^{D_{ai}(\theta - b_i)}} \qquad \text{(Eşitlik 6)}$$

Burada c_i parametresi düşük yetenek düzeyindeki yanıtlayıcıların maddeyi doğru yanıtlama olasılıklarını temsil etmektedir. Bu bakımdan şans başarısını ifade eden bir parametre olduğu belirtilmektedir. (Hambleton vd., 1991, s. 17; Hambleton ve Swaminathan, 1985, s. 37-38). c_i parametresinin yüksek olması durumunda, maddenin ayırt edicilik gücünün düşme eğiliminde olduğu belirtilmektedir. Bu nedenle maddenin sağladığı bilgi miktarının da düşmesine yol açmaktadır. Bir maddeye şansla doğru yanıt verme olasılığı arttıkça testi alan yanıtlayıcının θ yetenek düzeyini kestirmede sağladığı bilgi miktarı da daha az olmaktadır. Tersi durumunda ise şansla doğru yanıtlama olasılığı düştükçe yanıtlayıcının yetenek düzeyinin kestiriminde maddenin sağladığı kestirimdeki bilgi miktarının da arttığı ifade edilmektedir (Harvey ve Hammer, 1999).

1.2.4. 4 parametreli lojistik model (4plm)

4PLM, Barton ve Lord (1981)'un 3PLM'ye d_i parametresinin eklemesiyle geliştirilmiştir. Bu model ile yüksek yetenek düzeyindeki yanıtlayıcıların kolay bir maddeyi yanıtlamada hata yapabilme olasılıkları dikkate alınmaktadır. d_i parametresi, üst asimptotun 1,00'den küçük değerler alabilmesine olanak sağlamakta ve teorik olarak 0,00 ile 1,00 arasında değerler alabilmektedir. Değeri 1,00'den küçük olan üst asimptotun ilave edilmesi ile yüksek yetenek düzeyindeki bir yanıtlayıcının kolay bir maddeye yanlış yanıt vermesi durumunda yetenek ölçeğindeki yerinin önemli ölçüde değişmemesine olanak sağlamaktadır. Barton ve Lord (1981) 3PLM'nin, kolay bir maddeye yanlış yanıt veren yüksek yetenek düzeyindeki yanıtlayıcılar için gereğinden fazla cezalandırıcı olduğunu ifade etmişlerdir. 3PLM'de yetenek kestiriminde düşük yetenek düzeyindeki yanıtlayıcıların zor bir maddeyi şansla doğru yanıtlama olasılığı dikkate alınırken, 1,00 olan üst asimptot değeri ile yüksek yetenek düzeyindeki yanıtlayıcıların kolay bir maddeyi yanlış yanıtlama olasılıklarının sıfır olduğu kabul edilmektedir. 1,00'dan düşük d_i parametresi, bu olasılığın sıfırdan yüksek olması anlamına gelmekte ve böyle olduğunda, yüksek yetenek düzeyindeki yanıtlayıcılar söz konusu bu maddelere verdikleri yanlış yanıtlardan daha az puan kaybedeceklerdir. .

4PLM için bir yanıtlayıcının *i* maddesine doğru yanıt verme olasılığı Eşitlik 7'de sunulmaktadır:

$$P_i(\theta) = c_i + (d_i - c_i)\frac{e^{D_{ai}(\theta - b_i)}}{1 + e^{D_{ai}(\theta - b_i)}} \qquad \text{(Eşitlik 7)}$$

Diğer modellerde açıklanan parametrelere ek olarak bu eşitlikte d_i ile gösterilen üst asimptot, dikkatsizlik parametresi (inattention parameter) yer almaktadır. Genel anlamda 4PLM, her maddenin farklı üst asimptot değeri alabilmesine olanak sağlamaktadır. Fakat modelin ilk ortaya çıkarılma çalışmasında Barton ve Lord (1981) bütün maddeler için ortak bir üst asimptot değeri belirlemişler ve modeli o şekilde uygulamışlardır.

Barton ve Lord'un (1981) 4PLM konusundaki çalışmasında d_i parametresinin yetenek kestirimi konusunda a_i, b_i ve c_i parametrelerine ek olarak bir katkı sağlamadığı sonucuna ulaşılması, araştırmacıların bu modele ihtiyaç olmadığı yorumunu yapmalarına neden olmuştur. Hambleton ve Swaminthan (1985, s. 48) da kitabında Barton ve Lord'un çalışmasında vardığı bu sonuçlardan dolayı 4PLM'ye yeteri önem verilmemiş olabileceğini belirtmiştir. Bu doğrultuda 4PLM'ye 2000'li yıllara kadar önem verilmediği belirtilmektedir. Fakat o tarihten günümüze kadar bu model bazı araştırmacılar tarafından çalışılmıştır (Loken ve Rulison, 2010; Magis, 2013; Reise ve Waller, 2003; Rulison ve Loken, 2009; Rupp, 2003). Modelin çalışılmasında geç kalınmasına bir diğer sebep olarak, modelin geliştirilmesi sonrasında ortaya konan eserlerde, Barton ve Lord'un aslında 4PLM'ye ait *d* parametresini serbestçe kestirim yapmadığından bahsedilmediği ifade edilmektedir (Waller ve Reise, 2010). Bunun yerine Barton ve Lord'un çalışmasında üst asimptot değeri sırasıyla 1,00; 0,99; ve 0,98'e sabitlenerek analizler yapılmıştır.

1.3. Araştırmanın amacı

Bu araştırmada 1PLM, 2PLM, 3PLM ve 4PLM'nin 2012 SBS Türkçe alt testi verisine uyum düzeylerinin karşılaştırmalı olarak incelenmesi, bu modeller altında ilgili maddelerin ve alt testin toplam olarak sağladıkları bilgi miktarlarının hesaplanması ve model parametre kestirimlerinin doğruluğunun karşılaştırılması amaçlanmaktadır. Bu amaçla yapılan araştırmada aşağıdaki sorularına yanıt aranmıştır:

1. 1PLM, 2PLM, 3PLM ve 4PLM'den hangisi araştırma verisinde daha fazla uyum sağlamaktadır?
2. Hangi model altında madde parametreleri daha doğru bir şekilde kestirilmektedir?
3. Hangi model altında yetenek parametreleri (θ) daha doğru bir şekilde kestirilmektedir?
4. 1PLM, 2PLM, 3PLM ve 4PLM altında test bilgi fonksiyonuna dayalı olarak Türkçe alt testi hangi yetenek düzeyinde ne kadar bilgi vermektedir?

1.4. Araştırmanın önemi

1PLM, 2PLM, 3PLM'ye ek olarak Barton ve Lord (1981) 4PL modeli öne sürmüştür. 4PLM, yüksek yetenek düzeyindeki yanıtlayıcıların düşük güçlük düzeyine sahip maddeleri yanlış yanıtlama olasılıklarını kestiren *d* parametresini içermektedir. Magis (2013) de 4PLM altında kestirim yapıldığında madde ve test bilgi fonksiyonlarına ait değerlerin daha yüksek olduğunu ve dolayısıyla elde edilen bilgi miktarının fazla olduğunu belirtmektedir.

4PLM ile ilgili alan yazın incelendiğinde, modelin psikopatoloji alanında ve CAT uygulamalarında araştırıldığı görülmektedir (Loken ve Rulison, 2010; Reise ve Waller, 2003; Rulison ve Loken, 2009; Rupp, 2003; Waller ve Reise, 2010). Fakat Türkiye'de 4PLM ile ilgili bir çalışmaya rastlanmamaktadır. Oysa bu yeni modelin pratikteki önemi ve kullanılabileceği alanların araştırılması, bu konuda bilime katkı sağlamak ve uygulanabilirliğini arttırmak açısından önemli görülmektedir. Modele ait parametre kestirimlerinin açıklanması ve modelin, psikoloji alanının yanı sıra eğitimde ölçme alanındaki kullanılabilirliğinin ortaya konulması açısından incelemeye değer bulunmuştur.

Barton ve Lord (1981) 4PLM'ye dair yaptıkları çalışmada *d* parametresini sabitleyerek yetenek kestirimi yapmışlardır. Bu çalışma ise bir başarı testi üzerinde 1PLM, 2 PLM, 3PLM ve 4PLM ile analiz yapıldığından ve Barton ve Lord'un aksine parametre değerini sabitlemek yerine bütün parametrelerin

(*a, b, c, d*) serbestçe kestirileceğinden dolayı önemli görülmektedir. 4PLM altında madde ve yetenek parametre kestirimlerinin diğer modellere göre nasıl farklılaştığı, *d* parametresini dikkate almanın yetenek kestirimine katkı sağlayıp sağlamadığının incelenmesinin önemli olduğu düşünülmektedir. Yüksek yetenek düzeyindeki öğrencilerin, özellikle sonuçları yanıtlayıcının geleceği ile ilgili belirleyici olacak testlerde (high stakes) dikkatsizlik veya başka faktörler nedeniyle kolay maddeyi yanlış yanıtlama durumlarının, onların yetenek kestirimlerini etkilemesi ve bu kestirimlerle verilecek kararların doğruluğu açısından incelenmesinin önemli olduğu düşünülmektedir.

2. Yöntem

2.1. Araştırmanın modeli

Bu çalışmada, 2012 yılına ait Seviye Belirleme Sınav (SBS) Türkçe alt test maddelerinden elde edilen veri üzerinde 1PLM, 2PLM, 3PLM, 4PLM'ye dayalı olarak model veri uyumu, madde parametre kestirimlerinin doğruluğu, madde ve test bilgi fonksiyonları karşılaştırmalı olarak incelenmiştir. Bu amaç doğrultusunda bu çalışma, temel araştırma modelindedir. Temel araştırmalar, kuram geliştirmeyi ya da var olan kuramları sınamayı amaçlamaktadır. Kuramlara dayanarak, altında yatan süreçleri açıklayarak, bunları test ederek, sonuçlarını bilimsel olarak yorumlayarak bilgilerin genişlemesine katkı sunan araştırmalardır (Fraenkel, Wallen ve Hyun, 2012, s.7).

2.2. Çalışma grubu

2012 SBS'ye 1048575 ilköğretim kademesi sekizinci sınıf öğrencisi katılmıştır. Milli Eğitim Bakanlığı Ölçme, Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü tarafından bu araştırma için temin edilen 50000 yanıtlayıcıya ait veri setinden seçkisiz olarak seçilen 1500 yanıtlayıcı bu araştırmanın çalışma grubunu oluşturmaktadır.

2.3. Veri toplama araçları

2012 yılı SBS, 8. Sınıf öğrencilerine Türkçe, Matematik, Fen ve Teknoloji, Sosyal Bilgiler ve İngilizce alt testlerinde uygulanmıştır. Bu araştırmada, 23 maddeden oluşan 2012 yılı SBS Türkçe alt testinden elde edilen verilen kullanılmıştır

2.4. Verilerin çözümlenmesi ve yorumlanması

2.4.1. Varsayımların test edilmesi

2.4.1.1. Tek boyutluluk

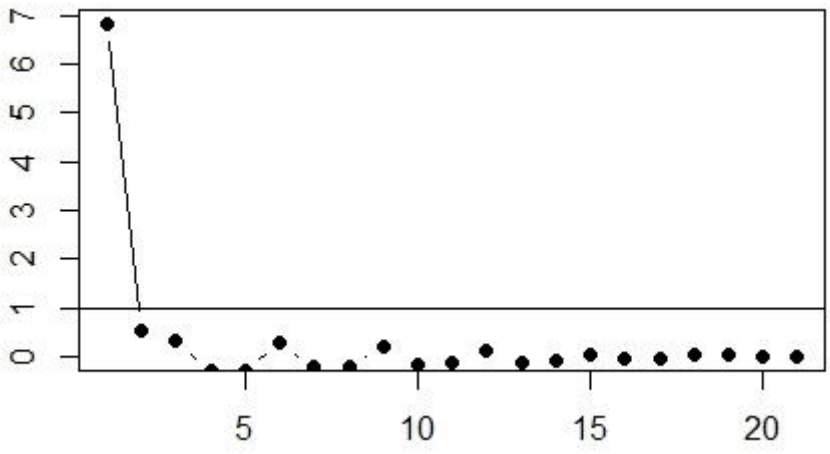
Bu çalışmada tek boyutluluğun test edilmesinde tetrakorik korelasyon matrisine dayalı açımlayıcı faktör analizinden (AFA) yararlanılmıştır. Bu amaç doğrultusunda R Studio programında 'psych' paketi kullanılmıştır (Revelle, 2015). AFA, 'maksimum likelihood' faktör çıkarma tekniği kullanılarak yapılmıştır. Veri setindeki bütün öğrenciler 23.maddeye doğru yanıt vermiştir ve dolayısıyla bu maddenin varyansı sıfır olduğu için program tarafından otomatik olarak analizden çıkarılmıştır.

Bu çalışmada faktör yük değerleri için sınır değer olarak 0,30 alınmıştır (Brown, 2015). Maddelere ilişkin faktör yük değerlerine bakıldığında tek bir maddeye ilişkin faktör yük değerinin 0,30'dan küçük olduğu görülmüştür (λ_{20} = 0,09). Bu doğrultuda 0,09 faktör yüküne sahip 20. maddenin ilgili faktörü yeterince temsil etmediği şeklinde yorumlanmış ve analizden çıkarılmıştır.

23. ve 20. maddeler çıkarıldıktan sonra yapılan faktör analizi sonucu hesaplanan faktör yükleri, 0,27 ila 0,77 arasında değişmektedir. Analiz sonucunda fit değeri 0,98 ve RMSEA indeksi 0,064 olarak bulunmuştur. Açıklanan varyans oranı %32 olup birinci faktöre ait özdeğer 6,82'dir. Bu bulgular testin tek faktörlü bir yapıya sahip olduğuna ve test maddelerinin ilgili özelliği tek boyutlu olarak ölçebildiğine işaret etmektedir. Faktör yükleri incelendiğinde bir madde hariç (m22) hepsinin 0,30'dan büyük faktör yüklerine sahip olduğu görülmektedir. 22. maddeye ait faktör yükü ise 0,27'dir ve alanyazında faktör

yük değeri 0,25'ten yüksek olan maddelerin ölçme aracında tutulmaya devam edildiği görülmektedir (Fleck, Poirier-Littre, Guelfi, Bourdel, ve Loo, 1995; Houben, Leeuw, Vlaeyen, Goubert, ve Picavet, 2005; Romera, Delgado-Cohen, Perez, Caballero, ve Gilaberte, 2008). Bu doğrultuda bilgi kaybına yol açmamak için bu maddenin testte tutulmasına karar verilmiştir.

Çalışma grubunu oluşturan 1500 yanıtlayıcıya ait 21 maddeden oluşan veri için yapılan faktör analizi sonucunda elde edilen özdeğerlere ilişkin eğri Şekil 1'de sunulmuştur. Test maddelerinin çoğuna ilişkin faktör yüklerinin 0,30'dan büyük olması ve özdeğeri 1'den büyük tek bir faktör bulunması ve grafikte görülen keskin düşüş, testin tek boyutluluk varsayımını karşıladığına işaret etmektedir.



Şekil 1 Testin 21 maddelik formuna ilişkin AFA sonucunda elde edilen özdeğerler eğrisi

2.4.1.2. Yerel bağımsızlık

Yerel bağımsızlık için alan yazında tek boyutluluk varsayımının karşılanmış olmasının, kanıt olarak sunulabildiği belirtilmektedir (Embretson ve Reise, 2000, s. 48; Hambleton vd., 1991, s. 11). Bu bilgiye ek olarak çeşitli istatistikler, yerel bağımsızlığın test edilmesinde kullanılmaktadır. Yen (1984), Q_3 istatistiğini yerel madde bağımsızlığı indeksi olarak ortaya koymuştur. Q_3 her bir madde çifti için hesaplanan artık değerlerin korelasyonudur. Hesaplanması için her bir yanıtlayıcı için yetenek kestirimi yapılır ve yanıtlayıcının her bir maddedeki beklenen performansını kestirmek için kullanılır. Artık değer, yanıtlayıcının bir maddedeki gözlenen ve beklenen performansı arasındaki sapma elde edilerek hesaplanır. Sonuçta seçilen iki maddenin Q_3 istatistiği, sapma puanlarının bütün yanıtlayıcılardaki korelasyonudur. Q_3 istatistiğinin değeri 0,20'den büyük olması ilgili madde çifti için yerel bağımsızlığı varsayımının sağlanamadığına işaret etmektedir (DeMars, 2010, s. 48). Yapılan çalışmalarda (Chen ve Thissen, 1997; Kim, De Ayala, Ferdous, ve Nering, 2007; Yen, 1993; Zenisky, Hambleton, ve Sireci, 2001) bu istatistik yerel bağımsızlığı test etmede başarılı bir şekilde kullanılmış ve yerel bağımlı olma durumunu ortaya çıkarmada diğer istatistiklere göre daha faydalı olduğu belirtilmiştir. Bu çalışmada, ilgili testte yer alan madde çiftleri için yerel bağımsızlık varsayımının karşılayıp karşılamadığının test edilmesi için her model altında bütün madde çiftlerine yönelik Q_3 istatistiği hesaplanmış ve elde edilen değerlerin 0,20'den büyük olup olmadığına bakılmıştır. Q_3 istatistiğinin hesaplanmasında R studio programı 'sirt' (Robitzsch, 2015) paketinden faydalanılmıştır.

Yerel bağımsızlık varsayımının test edilmesinde hesaplanan 21 maddeye ait 210 madde çiftine ilişkin Q_3 istatistikleri, her bir model altında ayrı ayrı hesaplanmıştır. Olası bütün madde çiftlerine ait

istatistiklerin bütün modeller altında 0,20'den küçük olduğu saptandığından ilgili veri seti için 1PLM, 2PLM, 3PLM ve 4PLM altında MTK yerel bağımsızlık varsayımının karşılandığı sonucuna ulaşılmıştır.

3. Bulgular

3.1. Birinci araştırma sorusuna ilişkin bulgular ve yorum

"1PLM, 2PLM, 3PLM ve 4PLM'den hangisi veriye daha fazla uyum sergilemektedir?" araştırma sorusuna yanıt vermek üzere öncelikle araştırma verisinin ilgili modellere genel uyum düzeyleri test edilmiş, ardından her bir model altında madde uyumu test edilmiştir.
Genel model uyumu için hesaplanan -2LL, AIC ve BIC değerleri ikili şekilde karşılaştırılmıştır. Tablo 1'de 1PLM ve 2PLM için hesaplanan -2LL, AIC ve BIC değerleri ile yapılan ANOVA sonucunda elde edilen p değeri ve serbestlik derecesi (sd) verilmiştir.

Tablo 1. 1PLM ve 2PLM için Model Veri Uyumu Karşılaştırması

	Model	-2LL	Parametre sayısı	AIC	BIC	sd	*p*
1	1PLM	17966,71	22	35977,41	36094,31	20	0
2	2PLM	17729,09	42	35542,19	35765,34		

Tablo 1 incelendiğinde, -2LL, AIC ve BIC değerlerinde 2PLM için azalma görülmektedir. Yapılan varyans analizi sonucunda $p = 0$ bulunmuştur. Bu bulgular, 2PLM'nin 1PLM'ye göre veriye daha iyi uyum sergilediğine işaret etmekte ve bu test verisi için 1PLM'ye göre daha kompleks olan 2PLM'nin kullanımının anlamlı ve gerekli olduğunu göstermektedir.

Diğer karşılaştırma 2PLM ve 3PLM arasında yapılmış ve elde edilen değerler Tablo 2'da sunulmuştur. 3PLM için hesaplanan -2LL ve AIC değerlerinin daha düşük olması ve manidar bulunan *p* değeri, 3PLM'nin 2PLM'ye göre daha iyi uyum gösterdiğine işaret etmektedir. BIC değerinde görülen bir miktar artış bu bulguyu desteklemese de varyans analizi sonucunun manidar çıkması 3PLM'nin 2PLM'ye tercih edilmesi gerektiğine işaret etmektedir.

Tablo 2. 2PLM ve 3PLM için Model Veri Uyumu Karşılaştırması

	Model	-2LL	Parametre sayısı	AIC	BIC	sd	*p*
1	2PLM	17729,09	42	35542,19	35765,34	21	0
2	3PLM	17681,66	63	35489,32	35824,05		

-2LL, AIC ve BIC değerlerine ilişkin son karşılaştırma 3PLM ve 4PLM arasında yapılmıştır. Tablo 3 incelendiğinde ikinci karşılaştırma (2PLM-3PLM, Tablo 8) ile benzer şekilde daha kompleks olan model (4PLM) için hesaplanan -2LL ve AIC değerlerinde düşüş, 3PLM ve 4PLM'ye ait BIC değerinde az miktarda artış görülmüştür. Varyans analizi ile yapılan karşılaştırmada *p* değeri manidar bulunmuştur. Bu, diğer karşılaştırmalarla benzer şekilde daha kompleks olan modelin test verisine daha uyumlu olduğunu göstermektedir. Bu bulgulara dayalı olarak araştırma verisinin en fazla 4PLM'ye uyum sergilediği sonucuna ulaşılmıştır.

Tablo 3. 3PLM ve 4PLM için Model Veri Uyumu Karşılaştırması

	Model	-2LL	Parametre sayısı	AIC	BIC	sd	*p*
1	3PLM	17681,66	63	35489,32	35824,05	21	0
2	4PLM	17620,39	84	35408,78	35855,09		

Araştırmada kullanılan bir diğer uyum istatistiği SRMSR değeridir. Bu değer sıfıra yakın olması ölçüsünde değerlendirilmektedir (Maydeu-Olivares vd., 2011). Tablo 4'te 1PLM, 2PLM, 3PLM ve 4PLM için SRMSR değerleri verilmiştir. Tablo incelendiğinde modellerdeki parametre sayısı arttıkça hata değerinin düştüğü ve sıfıra daha yakın olduğu görülmektedir. M_2 sınırlı bilgi uyum-iyiliği istatistiği için SRMSR değerleri, 4PLM'nin en az hata ile en çok uyumu gösterdiğine işaret etmektedir.

Tablo 1. 1PLM, 2PLM, 3PLM ve 4PLM için SRMSR Değerleri

	1PLM	2PLM	3PLM	4PLM
SRMSR	0,0743	0,0280	0,0255	0,0247

Son olarak ilgili 4 model için bütün karşılaştırma çiftleri dikkate alınarak LRT χ^2 hesaplanmış ve manidarlığı test edilmiştir. Tablo 11'de olası tüm model çiftleri için χ^2, sd ve *p* değerleri verilmiştir.

Tablo 5. LRT χ2 ile Model Veri Uyumu Karşılaştırması

	Model1	Model2	χ^2	sd	*p*
1	1PLM	2PLM	475,23	20	0
2	1PLM	3PLM	570,09	41	0
3	1PLM	4PLM	692,63	62	0
4	2PLM	3PLM	94,87	21	0
5	2PLM	4PLM	217,41	42	0
6	3PLM	4PLM	122,54	21	0

Tablo 5 incelendiğinde bütün model çiftleri için yapılan uyum iyiliği testinin manidar olduğu görülmektedir ($p = 0$). Bu bulgu her karşılaştırma için daha kompleks olan modelin daha iyi uyum gösterdiği ve eklenen parametrenin gerekli olduğunu göstermektedir. Yapılan genel model uyumu karşılaştırmalarına dayalı olarak, çalışma grubundaki öğrencilerin 2012 SBS Türkçe alt testinin 21 maddelik formunda yer alan maddelere yanıt verme davranışlarını en iyi açıklayan modelin 4PLM olduğu sonucuna ulaşılmıştır.

Model uyumu değerlendirmelerinde de kullanılan madde uyumu istatistikleri Bock'un (1972) χ^2 indeksi kullanılarak hesaplanmış ve sonuçlar Tablo 6'da verilmiştir. 1PLM, 2PLM ve 3PLM için ilgili test maddelerinin uyum değerleri sunulmuş ve $\alpha = 0{,}001$ manidarlık düzeyi esas alınmıştır. Madde uyumu için hesaplanan χ^2 değerinin manidar olması, maddenin uyum sergilemediğine işaret etmektedir.

Tablo 6. Madde Uyumu İndeksleri

	1PLM		2PLM		3PLM	
	χ^2	*p*	χ^2	*p*	χ^2	*p*
m1	39,4298	0,0003	35,3625	0,0085	26,9898	0,0582
m2	64,3116	**<0,0001**	33,7741	0,0134	25,8684	0,0769
m3	139,7372	**<0,0001**	51,1668	0,0001	35,5673	0,0052
m4	76,0226	**<0,0001**	38,6538	0,0032	29,1736	0,033
m5	184,961	**<0,0001**	29,5594	0,042	26,087	0,0729
m6	299,2595	**<0,0001**	44,3482	0,0005	55,1985	**<0,0001**
m7	72,5275	**<0,0001**	51,7398	**<0,0001**	33,3659	0,0101
m8	113,139	**<0,0001**	45,6418	0,0003	29,3744	0,0312
m9	295,2957	**<0,0001**	35,6646	0,0078	38,5111	0,0021
m10	122,266	**<0,0001**	44,6227	0,0005	48,7341	0,0001
m11	55,5634	**<0,0001**	26,4043	0,0909	21,4674	0,2061
m12	171,2558	**<0,0001**	47,5194	0,0002	41,5342	0,0008
m13	316,7562	**<0,0001**	31,4306	0,0257	34,0737	0,0082
m14	33,7286	0,0023	40,7102	0,0017	26,2003	0,0709
m15	240,7383	**<0,0001**	29,8766	0,0387	34,6179	0,007
m16	87,8219	**<0,0001**	31,5358	0,0249	28,6207	0,0382
m17	175,7137	**<0,0001**	26,2625	0,0939	28,4975	0,0395
m18	148,3522	**<0,0001**	37,0354	0,0052	38,3668	0,0022
m19	214,5814	**<0,0001**	51,7849	**<0,0001**	29,1803	0,0329
m21	10,8742	0,6959	16,6755	0,5455	33,2408	0,0105
m22	109,3191	**<0,0001**	35,3625	0,0085	26,8291	0,0606

Not: Manidar olan *p* değerleri kalın yazılmıştır.

Tablo 6 incelendiğinde 1PLM altında kestirim yapıldığında 1, 14 ve 21. Maddeler hariç diğer maddelerin uyum sergilemediği görülmektedir. 2PLM altında kestirim yapıldığında ise yalnız iki madde (7. ve 19.madde) uyum sergilememiştir. 3PLM için sadece bir maddenin (6.madde) uyum sergilemediği görülmektedir. Madde uyumu dikkate alındığında parametre sayısı arttıkça modele uyum sergilemeyen maddelerin azaldığı ve daha kompleks modellerin daha iyi uyum gösterdiği bulunmuştur. R studio programında ve başka herhangi bir programda 4PLM için madde uyumunu test eden bir kod bulunamadığı için 4PLM altında madde uyumu test edilememiştir. Bu nedenle tabloda görüldüğü gibi madde uyumu karşılaştırması yalnızca ilk 3 model için söz konusu olabilmiştir.

3.2. İkinci araştırma sorusuna ilişkin bulgular ve yorum

"Hangi model altında madde parametreleri daha doğru bir şekilde kestirilmektedir?" araştırma sorusuna yanıt vermek üzere 1PLM, 2PLM, 3PLM ve 4PLM altında kestirilen madde parametreleri ve bunlara ilişkin standart hatalar Tablo 7, Tablo 8, Tablo 9 ve Tablo 10'da sunulmuştur.

Tablo 7. 1PLM Altında Kestirilen Madde Parametreleri ve Standart Hatalar

Maddeler	***b***	**SH**
m1	0,21	0,06
m2	-0,87	0,07
m3	-0,34	0,06
m4	-0,54	0,06
m5	-1,10	0,07
m6	-1,55	0,08
m7	-0,50	0,06
m8	0,10	0,06
m9	-1,79	0,08
m10	-1,38	0,07
m11	-0,65	0,06
m12	0,32	0,06
m13	-1,18	0,07
m14	-0,64	0,06
m15	-1,19	0,07
m16	0,17	0,06
m17	-1,14	0,07
m18	-0,33	0,06
m19	-0,93	0,07
m21	-0,40	0,06
m22	0,74	0,06
Ortalama hata		0,07
	Ayırt edicilik	SH
	1,07	0,03

Tablo 7 incelendiğinde 1PLM altında maddelere ilişkin kestirilen *b* parametrelerinin -1,79 ile 0,74 arasında değiştiği, çoğunluğunun orta ve düşük düzeyde güçlük değerlerine sahip olduğu görülmektedir. En yüksek güçlük değerine sahip madde 22. maddedir. Güçlük değeri en düşük olan ise 9. maddedir. Her parametre kestirimi için ayrı ayrı hesaplanan standart hata değerleri kestirimin doğruluğunun derecesini vermektedir. *b* parametresine ilişkin standart hatalar 0,06 – 0,08 arasında değişmektedir. Bütün maddeler için hesaplanan *a* parametresinin değeri ise 1,07 ve hata oranı 0,03'tür. Bütün bu standart hata değerleri oldukça küçüktür ve bu bulgular, 1PLM altında madde parametrelerinin doğru bir şekilde kestirildiğine işaret etmektedir.

Tablo 8. 2PLM Altında Kestirilen Madde Parametreleri ve Standart Hatalar

Maddeler	*b*	SH	*a*	SH
m1	0,24	0,07	0,84	0,06
m2	-1,38	0,10	0,59	0,04
m3	-0,32	0,05	1,23	0,07
m4	-0,53	0,05	1,13	0,07
m5	-0,92	0,05	1,45	0,07
m6	-1,12	0,04	1,98	0,10
m7	-0,51	0,06	1,06	0,06
m8	0,13	0,08	0,72	0,06
m9	-1,25	0,04	2,00	0,11
m10	-1,33	0,06	1,14	0,06
m11	-0,73	0,06	0,91	0,06
m12	0,31	0,05	1,08	0,07
m13	-0,85	0,04	2,00	0,10
m14	-0,75	0,07	0,87	0,06
m15	-0,94	0,04	1,66	0,08
m16	0,23	0,08	0,67	0,06
m17	-0,98	0,05	1,39	0,07
m18	-0,29	0,05	1,35	0,08
m19	-0,78	0,04	1,47	0,08
m21	-0,47	0,06	0,87	0,06
m22	1,52	0,13	0,44	0,03
Ortalama hata		0,06		0,07

Tablo 8 incelendiğinde 2PLM altında maddelere ilişkin kestirilen güçlük parametrelerinin -1,38 ile 1,52 aralığında yer aldığı, *b* parametresinin kestirimine ilişkin standart hataların ise 0,04 – 0,13 aralığında değiştiği görülmektedir. Güçlük değeri en yüksek olan 22. madde iken en kolay madde ise 2. maddedir. 2PLM altında 0,44 ile 2,00 aralığında değerler alan ayırt edicilik parametresi incelendiğinde maddelerin genelinin yüksek ayırt ediciliğe sahip olduğu görülmektedir. Bu test için 9. ve 13. maddeler (*a*=2,00) en yüksek ayırt ediciliğe sahip maddelerdir. En düşük ayırt ediciliğe sahip madde ise 22. maddedir (*a*=0,44). Ayırt edicilik parametreleri kestirimine ilişkin standart hatalar ise 0,03–0,11 aralığında değişmektedir. Maddelere ilişkin *b* ve *a* parametresi standart hataları incelendiğinde ($\bar{X}_b$ = 0,06 ve $\bar{X}_a$ = 0,07) oldukça düşük değerlere sahip oldukları görülmektedir. Dolayısıyla 2PLM altında elde edilen madde parametrelerinin doğru bir şekilde kestirildiği yorumu yapılabilir.

Tablo 9. 3PLM Altında Kestirilen Madde Parametreleri ve Standart Hatalar

Maddeler	*b*	SH	*a*	SH	*c*	SH
m1	0,62	0,06	1,17	0,10	0,14	0,02
m2	0,66	0,08	1,77	0,23	0,53	0,02
m3	0,15	0,04	1,93	0,14	0,21	0,02
m4	0,00	0,05	1,56	0,12	0,22	0,02
m5	-0,56	0,05	1,73	0,11	0,19	0,03
m6	-1,14	0,04	1,89	0,09	0,00	0,03
m7	0,19	0,05	1,73	0,14	0,28	0,02
m8	0,85	0,06	1,56	0,15	0,27	0,02
m9	-1,23	0,04	2,00	0,10	0,05	0,04
m10	-1,35	0,06	1,11	0,05	0,00	0,04
m11	-0,23	0,06	1,12	0,08	0,19	0,02
m12	0,59	0,05	1,58	0,12	0,13	0,02
m13	-0,79	0,04	2,00	0,10	0,05	0,02
m14	-0,75	0,07	0,86	0,05	0,00	0,03
m15	-0,72	0,04	1,83	0,10	0,13	0,03
m16	0,87	0,08	1,10	0,10	0,22	0,02
m17	-0,80	0,05	1,47	0,08	0,10	0,03
m18	-0,04	0,04	1,64	0,11	0,11	0,02
m19	-0,27	0,04	2,00	0,14	0,26	0,02
m21	-0,46	0,07	0,86	0,06	0,00	0,03
m22	1,52	0,13	0,44	0,03	0,00	0,02
Ortalama hata		0,06		0,11		0,02

Tablo 9 incelendiğinde 3PLM altında maddelere ilişkin kestirilen güçlük parametrelerinin -1,35 ile 1,52 aralığında değerler aldığı, *b* parametresinin kestirimine ilişkin standart hataların ise 0,04 – 0,13 aralığında değiştiği görülmektedir. Ortalama standart hatası 0,06 olan *b* parametrelerinin düşük hata düzeyi ile bu model altında doğru bir şekilde kestirildiği görülmüştür. 1PLM ve 2PLM ile benzer şekilde 3PLM altında kestirim yapıldığında en zor maddenin yine 22. madde olduğu gözlenmiştir. En düşük güçlük düzeyine sahip madde ise 10. maddedir. 3PLM altında 0,44 ile 2,00 aralığında değerler alan *a* parametresi incelendiğinde maddelerin ayırt edicilik düzeylerinin yüksek olduğu görülmektedir. Ayırt edicilik düzeyi en yüksek olan maddeler 9, 13 ve 19. maddelerdir. Yine 2PLM ile benzer şekilde en düşük ayırt ediciliğe sahip madde 22. maddedir. *a* parametresi kestirimlerine ilişkin standart hatalar ise 0,03 – 0,23 aralığında değişmektedir. Ortalama hata değeri $\bar{X}_a$ = 0,11 ile diğer parametrelere göre nispeten yüksek olan *a* parametresi kestirimleri de çok yüksek hata değerlerine sahip olmadığı için doğruluk derecesi kabul edilebilir niteliktedir. *a* parametresi kestirimi açısından standart hatanın en yüksek olduğu madde 2. maddedir (0,23). Bu madde ayırt ediciliği yüksek bir maddedir ve hata oranı yüksek olduğu için ne kadar ayırt edici olduğunun kestirilmesi zordur.

c parametre değerlerine bakıldığında birçok maddenin sıfırdan farklı değerler aldığı görülmektedir ve bu maddeler için değerler 0,05 ile 0,53 aralığında değişmektedir. Özellikle 2. madde 0,53 *c* parametre değeri ile şansla doğru yanıtlanma olasılığı en yüksek madde olarak dikkat çekmektedir. *c* parametrelerine ilişkin standart hatalar ise 0,02 ile 0,04 aralığında değişmektedir. Ortalama standart hata değeri $\bar{X}_c$=0,02 olan *c* parametrelerinin doğru bir şekilde kestirildiği görülmektedir. Sıfırdan farklı *c* parametre değerlerine sahip maddelerin çokluğu da model uyumundaki bulgu ile paralel bir şekilde şans parametresinin dâhil edildiği 3PLM'nin kullanımının 2PLM'ye göre gerekli olduğuna işaret etmektedir.

Tablo 10. 4PLM Altında Kestirilen Madde Parametreleri ve Standart Hataları

Maddeler	*b*	SH	*a*	SH	*c*	SH	*d*	SH
m1	0,19	0,06	2,00	0,24	0,19	0,02	0,78	0,02
m2	0,57	0,08	2,00	0,30	0,53	0,02	0,97	0,02
m3	0,02	0,05	2,00	0,17	0,20	0,02	0,96	0,01
m4	-0,09	0,05	2,00	0,19	0,25	0,02	0,94	0,01
m5	-0,60	0,05	2,00	0,15	0,21	0,03	0,97	0,01
m6	-1,18	0,04	2,00	0,11	0,00	0,03	0,99	0,01
m7	0,16	0,05	1,68	0,14	0,27	0,02	1,00	0,01
m8	0,63	0,06	2,00	0,24	0,28	0,02	0,89	0,02
m9	-1,26	0,04	2,00	0,10	0,04	0,04	1,00	0,00
m10	-1,29	0,05	2,00	0,16	0,05	0,04	0,92	0,01
m11	-0,30	0,06	1,08	0,08	0,17	0,02	1,00	0,02
m12	0,33	0,05	2,00	0,19	0,14	0,02	0,86	0,02
m13	-0,81	0,04	2,00	0,10	0,04	0,02	1,00	0,00
m14	-1,03	0,07	1,19	0,09	0,00	0,03	0,88	0,01
m15	-0,73	0,04	2,00	0,13	0,15	0,03	0,99	0,01
m16	0,29	0,07	2,00	0,29	0,25	0,02	0,76	0,02
m17	-0,71	0,05	2,00	0,16	0,19	0,03	0,96	0,01
m18	-0,07	0,04	1,71	0,12	0,12	0,02	0,99	0,01
m19	-0,29	0,04	2,00	0,14	0,26	0,02	1,00	0,01
m21	-0,42	0,06	2,00	0,23	0,21	0,02	0,82	0,02
m22	-0,95	0,12	1,00	0,14	0,00	0,03	0,50	0,02
Ortalama hata		0,06		0,17		0,02		0,01

Tablo 10 incelendiğinde 4PLM altında maddelere ilişkin kestirilen *b* parametrelerinin -1,29 ile 0,63 arasında değerler aldığı ve bunlara ilişkin standart hataların 0,04 ile 0,12 aralığında olduğu görülmektedir. En kolay madde 10. madde (b_{10}= 1,29) ve en zor madde 8. maddedir (b_8=0,63). Madde güçlük parametre kestirimlerinin 3PLM'ye kıyasla daha düşük olduğu görülmektedir. Bu model altında kestirilen *a* parametrelerinin 1,00 ile 2,00 arasında değerler aldığı ve büyük çoğunluğunun (16 madde) *a*=2,00 parametre değeri ile ayırt edicilik düzeylerinin çok yüksek olduğu görülmektedir. Bu bulgu, *d* parametresinin dikkate alındığı 4PLM altında kestirim yapıldığında maddelerin yüksek ve düşük yetenekteki bireyleri daha iyi ayırdığını göstermektedir. Fakat a parametresine ilişkin ortalama hatanın nispeten yüksek olduğu söylenebilir.

4PLM altında kestirilen *c* parametre değerleri incelendiğinde 3PLM ile benzer şekilde birçok maddenin sıfırdan farklı değerler aldığı görülmektedir. Elde edilen sıfırdan farklı *c* parametre değerleri 0,04 ile 0,53 aralığında değişmektedir. 3PLM'de olduğu gibi yine en yüksek *c* parametre değerine sahip maddenin 2. madde (c_2= ,53) olduğu görülmüştür. *c* parametrelerinin standart hataları ise 0,02 ile 0,04 aralığında değişmektedir ve oldukça küçüktür. Küçük standart hata değerleri *c* parametrelerine ilişkin kestirimin doğruluk derecesinin yüksek olduğunu göstermektedir.

Tablo 10'daki *d* parametre kestirimleri incelendiğinde 17 maddenin 1,00'dan farklı değerler aldığı görülmektedir. İlgili maddeler için *d* parametre kestirimleri 0,50 ile 1,00 arasında değişmektedir. Dikkatsizlik sebebiyle yanlış yanıtlanma olasılığı en yüksek madde 0,50 ile 22. Maddedir. Maddeye ait güçlük parametresi -0,95'tir ve kolay bir madde olduğu dikkate alındığında, bu maddeye ilişkin *d* parametresinin 1.00'dan oldukça düşük olması, yüksek yetenek düzeyindeki yanıtlayıcıların bu maddeyi dikkatsizlik ve benzeri nedenlerle yanlış yanıtlama olasılıklarının çok yüksek olduğuna işaret etmektedir. Ayrıca dört madde için *d* parametresi 1,00 olarak kestirilmiştir. Bu, *d* parametre kestiriminin madde temelli olduğunu ve bütün maddelere sabit bir değer atamak yerine serbestçe kestirim yapılması gerektiğini göstermektedir. *d* parametresinin kestirimine ilişkin standart hata değerleri ortalaması $\bar{X}_d$ = 0,02 olup, oldukça düşüktür ve kestirimin doğru bir şekilde yapıldığına işaret etmektedir. İlgili birçok

madde için elde edilen 1.00'dan farklı *d* parametre kestirimi ilgili test için 4PLM'e ihtiyaç duyulduğunu göstermektedir.

3.3. Üçüncü araştırma sorusuna ilişkin bulgular ve yorum

"Hangi model altında yetenek parametresi (θ) daha doğru bir şekilde kestirilmektedir?" araştırma sorusuna yanıt vermek üzere çalışma grubundaki öğrenciler için 1PLM, 2PLM, 3PLM ve 4PLM altında yetenek kestirimi yapılmış ve bu kestirimlere ilişkin standart hatalar hesaplanmıştır. Kestirimin doğruluğu konusunda bilgi veren standart hata değerlerine ait betimsel istatistikler Tablo 17'de sunulmuştur. En yüksek hata ortalamasına sahip model 2PLM'dir ($\overline{X}$ =0,54). 3PLM ve 4PLM için hata ortalamalarının giderek düştüğü, en düşük hata ortalamasına sahip modelin 4PLM ($\overline{X}$ =0,41) olduğu görülmektedir.

Tablo 11. Yetenek Kestirimi Standart Hatalarının Betimsel İstatistikleri

Modeller	N	$\overline{X}$	SS	Çarpıklık	Basıklık
1PLM	1500	0,52	0,13	2,21	4,85
2PLM	1500	0,54	0,25	2,60	8,47
3PLM	1500	0,49	0,18	3,70	17,05
4PLM	1500	0,41	0,12	3,60	17,45

İlgili dört model altında hesaplanan hata ortalamaları arasında manidar düzeyde bir farklılık olup olmadığının belirlenmesi amacıyla yapılan ilişkili örneklemler için tek faktörlü Varyans Analizi (ANOVA) sonucu Tablo 12'de sunulmuştur.

Tablo 2. Farklı Modeller Altındaki Yetenek Kestiriminin Standart Hatalarına Ait ANOVA Sonuçları

Varyans Kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	*p*	Manidar Fark
Modeller	14,49	3	4,83	431,25	0,000	4PLM<3PLM<1PLM<2PLM
Hata	50,37	4497	0,11			

Analiz sonuçları, farklı modeller (1PLM, 2PLM, 3PLM ve 4PLM) altındaki yetenek kestirimlerine ilişkin standart hata değerleri arasında manidar düzeyde bir farklılık olduğunu göstermektedir ($F_{3,4497}$=431,25; p<0,05). Bu bulgu, bireylere ilişkin yetenek kestiriminin doğruluğunun seçilen modele göre değiştiğini göstermektedir. Gözlenen bu farklılığın, hangi modellere ilişkin standart hatalar arasında olduğunu araştırmak için yapılan ikili karşılaştırmalar, olası bütün model karşılaştırmaları arasında manidar düzeyde farklılıkların olduğunu göstermiştir. Dolayısıyla Tablo 11'deki standart hata ortalamaları dikkate alındığında gözlenen bu farklılığın, 2PLM'den itibaren birbirine göre daha kompleks olan model lehine olduğu görülmektedir. Hata ortalamaları modellerdeki parametre sayısı arttıkça azalmaktadır. Yani 3PLM altında kestirilen yeteneklerin standart hatası, 2PLM altında kestirilen yeteneklerin standart hatasından ve 4PLM altında kestirilen yeteneklerin standart hatası 3PLM altında kestirilen yeteneklerin standart hatasından manidar düzeyde daha düşüktür. 4PLM altındaki yetenek kestirimine ilişkin hata değerlerinin, diğer üç modelden manidar olarak daha düşük olduğu görülmektedir. Manidar farklılıklara ilişkin etki büyüklükleri yalnızca sıralı modeller için hesaplanmış ve Tablo 13'de sunulmuştur.

Tablo 3. İkili Karşılaştırmalardaki Manidar Farklara İlişkin Etki Büyüklüğü η^2

	2PLM – 1PLM	3PLM – 2PLM	4PLM – 3PLM
η^2	0,02	0,09	0,32

Tablo 13'e bakıldığında, 2PLM- 1PLM karşılaştırması ile 3PLM-2PLM karşılaştırmasına ilişkin hesaplanan η^2 değerlerinin çok küçük olduğu görülmektedir. Bu bulgu, ilgili modeller için, modelin yetenek parametresinin kestiriminin doğruluğundaki değişkenliği çok az düzeyde açıklayabildiğine işaret etmektedir. Başka bir ifadeyle ilgili modeller için, modelin yetenek kestirimimin

doğruluğu üzerinde çok az bir etkisi bulunmaktadır. 4PLM'ye ve 3PLM'ye dayalı yetenek kestirimlerinin doğruluğuna ilişkin karşılaştırma için etki büyüklüğü $\eta^2 = 0,32$ bulunmuştur. Bu değer, ilgili modeller için, modelin yetenek parametresi kestirimlerinin doğruluk düzeylerindeki değişimi orta düzeyde açıklayabildiğine, modelin (3PLM ve 4PLM için) yetenek kestiriminin doğruluğu üzerinde orta düzeyde bir etkiye sahip olduğuna işaret etmektedir.

Yetenek kestirimine ilişkin standart hataların karşılaştırılmasına yönelik bu bulgular, 2PLM'den itibaren modellerdeki parametre sayısı arttıkça yetenek kestirimine ilişkin standart hata ortalamalarının manidar bir şekilde azaldığını ve 4PLM'nin en düşük hata ortalamasına (0,41) sahip olduğunu göstermiştir. ANOVA sonuçları ve etki büyüklüğü indeksleri değerlendirildiğinde, *d* parametresinin dâhil edildiği 4PLM altında kestirim yapıldığında, çalışma grubundaki öğrencilerin yetenek düzeylerinin en az hata ile kestirildiği sonucuna ulaşılmıştır.

3.4. Dördüncü araştırma sorusuna ilişkin bulgular ve yorum

"1PLM, 2PLM, 3PLM ve 4PLM altında test bilgi fonksiyonuna dayalı olarak Türkçe alt testi hangi yetenek düzeyinde ne kadar bilgi vermektedir?" araştırma sorusuna yanıt vermek üzere öncelikle madde ve test bilgi fonksiyonları hesaplanmıştır.

Test bilgi fonksiyonu, madde bilgi fonksiyonlarının toplamı olarak ifade edilmektedir. Bu nedenle öncelikle ilgili maddelerin $\theta = -4$ ile $\theta = 4$ yetenek aralığında verdiği bilgi miktarları her bir model altında ayrı ayrı hesaplanmış ve Tablo 14'te sunulmuştur.

Tablo 4. θ Yetenek Ölçeği (θ = [-4] – [4]) Üzerinde Maddelerin Sağladıkları Toplam Bilgi Miktarları

	1plm	2plm	3plm	4plm
m1	12,79	9,61	9,42	14,92
m2	12,65	5,68	6,20	6,91
m3	12,78	14,82	13,94	14,73
m4	12,74	13,44	10,89	13,09
m5	12,55	17,55	12,95	14,43
m6	12,26	24,14	23,03	24,42
m7	12,75	12,51	10,68	10,54
m8	12,80	7,83	9,79	12,33
m9	12,04	26,09	20,63	21,39
m10	12,38	13,33	12,95	20,63
m11	12,72	10,52	8,27	8,36
m12	12,78	12,89	13,36	16,57
m13	12,51	25,20	20,80	21,27
m14	12,72	9,95	9,71	14,12
m15	12,50	20,24	15,53	16,15
m16	12,79	7,11	7,52	13,06
m17	12,53	16,81	13,32	14,92
m18	12,78	16,41	14,43	14,88
m19	12,63	17,79	13,04	13,04
m21	12,77	10,04	9,80	14,45
m22	12,69	3,61	3,64	11,69

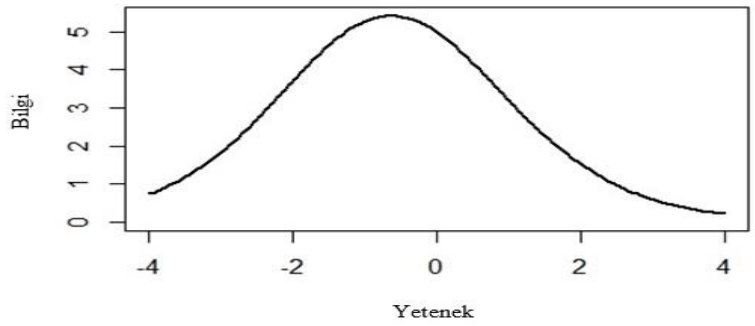
Tablo 14 incelendiğinde 1PLM altında ilgili 21 maddenin hepsinin hemen hemen birbirine yakın değerler aldığı ve çeşitlilik göstermediği görülmektedir. Ayırt edicilik bilgi fonksiyonunu etkilemektedir. 1PLM için tüm maddelerde ayırt edicilik eşit olduğu için madde bilgi miktarları birbirine çok yakın değerler almıştır. Sekiz madde için (1, 2, 8, 12, 14, 16, 21 ve 22) 2PLM'den 4PLM'ye gittikçe bilgi miktarında artış gözlenmiştir. Tablo 16'daki *d* parametre kestirimlerine bakıldığında, *d* parametre kestirimi 1,00'dan farklı ($d_1 = 0,78$; $d_2 = 0,97$; $d_8 = 0,89$; $d_{12} = 0,89$; $d_{14} = 0,88$; $d_{16} = 0,76$; $d_{21} = 0,82$;

d_{22} = 0,50) olduğu gözlenen bu maddeler için 4PLM altında analiz yapıldığında maddelerin daha fazla bilgi verdiği görülmektedir. İlgili dört model için test bilgi miktarları bu çalışmada esas alınan θ = -4 ile θ = 4 aralığı için ve bu aralığın içerisindeki sekiz düzey için hesaplanmış, Tablo 15'te sunulmuştur.

Tablo 15. θ Yetenek Düzeylerine Göre Test Bilgi Miktarları

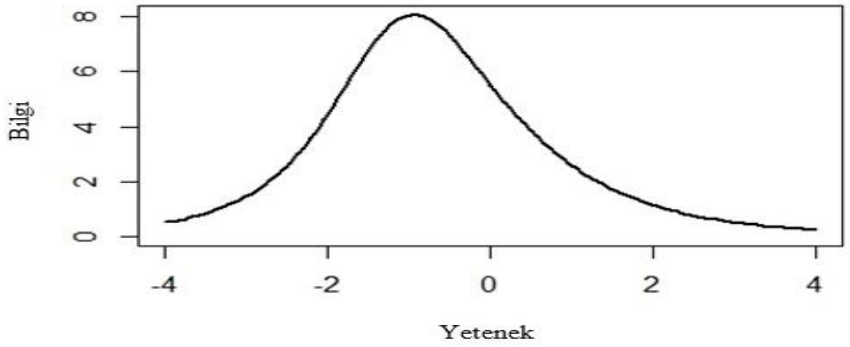
	(-4,00) (-3,10)	**(-3,02) (-2,12)**	**(-2,04) (-1,14)**	**(-1,06) (-0,08)**	**(0,08) (1,06)**	**(1,14) (2,04)**	**(2,12) (3,02)**	**(3,10) (4,00)**	**(-4)(4) Toplam**
1plm	13,89	31,01	53,25	73,99	52,07	25,62	11,08	4,23	265,13
2plm	10,14	29,91	74,83	100,10	48,17	19,61	8,79	4,04	295,58
3plm	3,34	11,98	42,82	89,44	69,06	29,70	10,18	3,38	259,90
4plm	2,45	12,74	53,25	119,62	90,17	27,11	5,49	1,08	311,91

Tablo 15 incelendiğinde ilgili bütün modellerin orta yetenek düzeylerinde daha çok bilgi sağladığı ve hepsinin ortak olarak en çok bilgi sağladığı yetenek düzeyinin θ = -1,06 ile θ = -0,08 aralığında olduğu görülmektedir. Uçlara gidildikçe bilgi miktarı azalmaktadır. SBS'nin bir seviye belirleme sınavı olduğu dikkate alındığında maddelerin orta yetenek düzeyindeki öğrencilere hitap etmesi önemli görülmektedir. Özellikle -1,06 ile -0,08 ve 0,08 ile 1,06 yetenek düzeyleri için en çok bilgi (sırasıyla; 119,62 ve 90,17) sağlayan modelin 4PLM olması dikkat çekici bir bulgudur. θ = -4 ile θ = 4 yetenek aralığı için toplam bilgi miktarı incelendiğinde ilgili 21 maddenin 4PLM altında analizi sonucunda en çok bilgi (311,91) sağladığı görülmektedir. Bu bulgu, 4PLM'nin ilgili diğer modellere göre bireyin yeteneğinin kestirimi konusunda daha fazla bilgi sağladığını göstermektedir. Yapılan bu analiz sonucundaki toplam test bilgi miktarı en çok olan ve orta yetenekte en çok bilgiyi sağlayan modelin (4PLM) kullanımının bu test için faydalı olabileceği şeklinde yorumlanabilmektedir.



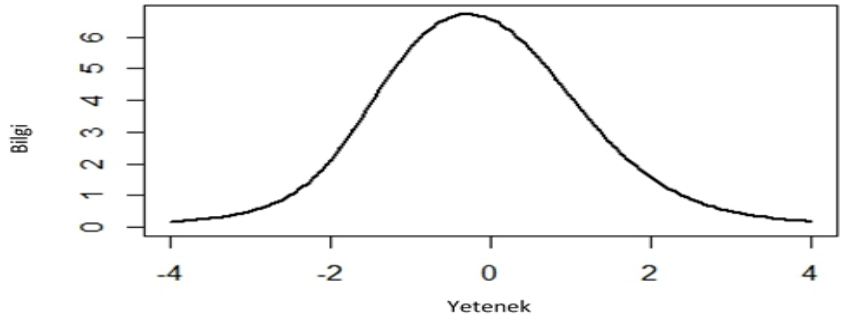
Şekil 2 1PLM için test bilgi fonksiyonu

1PLM altında ilgili 21 maddeden oluşan teste ait test bilgi fonksiyonu grafiği Şekil 2'de sunulmuştur. 1PLM için hesaplanan test bilgi fonksiyonu grafiğinin diğer modellere göre daha yatay olduğu gözlenmektedir. En çok bilgi verdiği yetenek kestirimi aralığı θ = -1 ve çevresidir. Şekil 3'te ise ilgili teste ait 2PLM altındaki test bilgi fonksiyonu sunulmuştur.



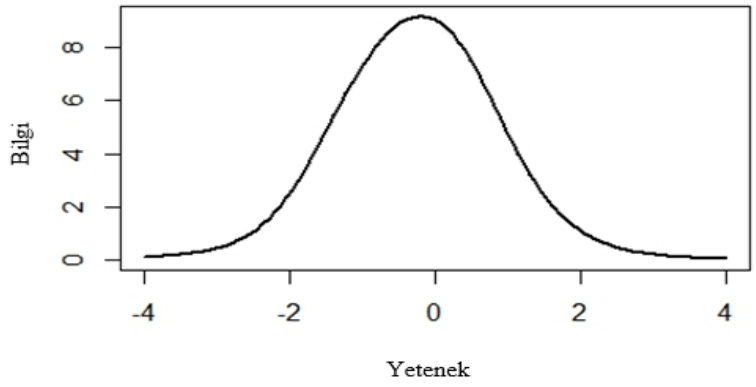
Şekil 3 2PLM için test bilgi fonksiyonu

Şekil 3 incelendiğinde 1PLM'den farklı olarak daha dik bir eğri görülmektedir. Fakat benzer şekilde 2PLM altında kestirim yapıldığında ilgili test maddelerinin toplam olarak en çok bilgi verdiği yetenek kestirimi aralığı θ = -1 ve çevresidir. 3PLM'ye ait test bilgi fonksiyonu grafiği ise Şekil 4'te sunulmuştur.



Şekil 4 3PLM için test bilgi fonksiyonu

Şekil 4 incelendiğinde 3PLM altında kestirim yapıldığında ilgili testin θ = 0 ve çevresi için daha fazla bilgi sağladığı görülmüştür. Şekil 5'te yer alan 4PLM altındaki test bilgi fonksiyonu da benzer şekilde θ = 0 ve çevresindeki yanıtlayıcılar için testin daha fazla bilgi sağladığını göstermektedir. Yetenek ölçeğinde uçlara gidildikçe testin sağladığı bilgi miktarı azalmaktadır. Grafikler de bilgi miktarlarındaki bulguyu destekler nitelikte olup 4PLM'nin orta yetenek düzeyinde diğer modellere göre daha fazla bilgi sağladığını göstermektedir.



Şekil 5 4PLM için test bilgi fonksiyonu

Bilgi fonksiyonları yetenek kestiriminin doğruluğu hakkında da bilgi vermektedir. Bilgi fonksiyonu yetenek kestiriminin standart hatası ile ters orantılıdır. Bilgi miktarı ne kadar arttıkça standart hata düşmektedir. 4PLM altında kestirim yapıldığında en çok bilgi miktarı elde edildiği dikkate alındığında yetenek kestirimi standart hatasının da en düşük bu model altında olduğu söylenebilir.

4. Sonuç ve tartışma

Bu araştırmada, 2012 SBS Türkçe alt testinden elde edilen verinin 1PLM, 2PLM, 3PLM ve 4PLM'den hangisine daha iyi uyum sergilediği, madde ve yetenek parametre kestirimlerinin hangi model altında daha doğru bir şekilde kestirildiği ve testin hangi model altında daha fazla bilgi verdiği incelenmiştir. Bu amaç doğrultusunda ilgili test için genel uyum ve madde uyum indeksleri hesaplanmış, madde ve yetenek parametreleri kestirilmiş ve test bilgi fonksiyonları elde edilmiştir. Analizlere geçilmeden önce MTK'nın varsayımları olan tek boyutluluk ve yerel bağımsızlığın karşılanıp

karşılanmadığı incelenmiş ve varsayımların bütün modeller için karşılandığı sonucuna ulaşılmıştır. Yapılan AFA sonucunda testin tek boyutlu olduğu sonucuna ulaşılmış ve geçerlik kanıtı olarak sunulmuştur.

Genel model veri uyumu 5 farklı yöntemle (-2LL, AIC, BIC, χ^2 ve SMRSR) sınanmıştır. Yapılan karşılaştırmalarda kullanılan yöntemlerden dördü 4PLM'nin göreli olarak daha iyi uyum sergilediğine işaret etmektedir. 2PLM için hesaplanan BIC değeri ise diğer modellere ait BIC değerlerinden daha düşük değer almış ve bu 2PLM'nin daha iyi uyum sergilediğine işaret etmiştir. Benzer şekilde Kang, Cohen ve Sung (2005) da çalışmalarında AIC ve BIC değerlerinin genelde log-likelihood ile yapılan karşılaştırma ile benzer sonuçlar sunduğunu fakat test maddelerinin kolay olduğu durumlarda, özellikle BIC değerinin 2 parametreli lojistik modelin daha iyi uyum sergilediği yönünde sonuçlar üretmeye meyilli olduğunu belirtmişlerdir. Madde uyumu ilk üç model için test edilebilmiş ve 3PLM altında kestirim yapıldığında neredeyse bütün maddelerin uyumlu olduğu görülmüştür. Barton ve Lord (1981), GRE, SAT ve AP gibi geniş ölçekli testleri kullandıkları çalışmalarında likelihood değerleri dikkate alındığında 4PLM lehine bir fark olmadığını belirtmişlerdir. Fakat bu çalışmada *d* parametresi serbestçe kestirilmemiş, bütün maddeler için tek bir *d* parametre kestirimi sabitlenerek karşılaştırma yapılmıştır. Loken ve Rulison (2010) da 4PLM altında serbestçe parametre kestirimi yapmış ve model uyumunu değerlendirmişlerdir. Simülasyon ve gerçek veri seti ile yaptıkları çalışmada 4PLM'nin 2PLM ve 3PLM'ye göre daha iyi uyum sağladığını bulmuşlardır.

Testlerin, amacına göre farklı yetenek düzeylerine hitap etmesi beklenebilir. İstenilen bu düzeyler için diğerlerine göre daha doğru ölçümler elde etmek istenebilir. Madde güçlük parametreleri, testin daha doğru bir şekilde ölçülmesinin istendiği yetenek düzeyini hedef alıp almadığını değerlendirmede kullanılabilir. Çoğu başarı testinde (seviye belirleme gibi) her yanıtlayıcının yetenek düzeyi en doğru şekilde ölçülmek istenmektedir. Bu nedenle bu testlerde *b* parametresi kestirimlerinin geniş aralıkta olması beklenmektedir. Kesme puanının olduğu durumlarda, maddelerin güçlük düzeylerinin, belirlenen yetenek düzeyinde yoğunlaşması ve o yetenek aralığı/noktası için daha doğru yetenek kestirimi yapılması istenmektedir. Bu araştırmada incelenen seviye belirlemeye yönelik test maddelerine ilişkin *b* parametre kestirimlerinin dar bir aralıkta olmamakla birlikte istenilen genişlikte de olmadığı, çoğunun orta güçlükte olması beklenirken kolay maddelerin çok olduğu görülmüştür. Güçlük parametresinin kestirimine ilişkin standart hatalar bütün modeller için de oldukça düşüktür ve parametre kestirimlerinin doğruluğuna işaret etmektedir.

Ayırt edicilik parametresi, maddenin yüksek ve düşük yetenekteki bireyleri ne ölçüde ayırdığını göstermektedir. Bu araştırmadaki ilgili test maddeleri dört model için de yüksek ayırt ediciliğe sahiptir. Modellerdeki parametre sayısı arttıkça ayırt edicilik değerlerinin arttığı ve en yüksek değerlere 4PLM altında ulaşıldığı görülmüştür. Bir madde için, madde parametrelerinin birbirinden bağımsız bir şekilde kestirilmediği, kestirimlerin ya da eklenen parametrelerin birbirini etkilediği belirtilmektedir. Örneğin *c* parametresi dikkate alınmadığında *b* parametresi de olduğundan daha düşük kestirilecektir (şans başarısına dikkat edilmediği için madde daha kolay görünecektir) (DeMars, 2010, s. 66). *d* parametresi dikkate alındığında madde karakteristik eğrileri üst asimptota yaklaşan daha dik/keskin bir biçimde şekillenmektedir. Bu durum aynı zamanda eğrilerin yükselmesine de neden olmaktadır. Eğimin dik olması *a* parametre kestiriminin yüksek olmasına, yüksekliği de *b* parametre kestiriminin düşük olmasına işaret etmektedir. 4PLM altında kestirim yapıldığında ve madde karakteristik eğrisi üst asimptot değeri 1,00'dan küçük olduğunda, maddeye ilişkin ayırt edicilik artmakta ve güçlük azalmaktadır (Loken ve Rulison, 2010). Bir başka ifadeyle *d* parametresi 1,00'dan farklı olan maddeler için 4PLM altında kestirim yapılması, yüksek yetenek düzeyindeki yanıtlayıcıların o maddeye doğru yanıt verme olasılıklarının artması anlamına gelmekte ve bu durum da maddenin ayırt ediciliğini arttırmakta fakat güçlüğünü azaltmaktadır. Benzer şekilde Waller ve Reise (2010) çalışmalarında, 4PLM altında kestirim yapıldığında, 3PLM'ye kıyasla *a* parametre kestirimlerinin arttığını ve *b* parametre kestirimlerinin düştüğünü ifade etmektedirler. Çalışmada bu durumun temel geometrik bilgiler ışığında

açıklanabileceği belirtilmiştir: Alt asimptot 0,00'dan büyük ve üst asimptot 1,00'dan küçük olduğu için madde karakteristik eğrisinin bu dar yetenek aralığı içerisinde hızlı bir şekilde yükselmesi gerektiği belirtilmektedir. Eğrideki bu hızlı yükseliş de ayırt ediciliğin yüksek olmasına işaret etmektedir.

Ayırt edicilik parametrelerine ilişkin standart hataların da modeller daha kompleks hale geldikçe arttığı görülmüştür. Normalin dışında *a* parametreleri elde edilmesi sebebiyle kestirim sırasında *a* parametresi için sınırlandırmaya gidilmiş olması hata değerlerindeki bu artışın sebebi olarak gösterilebilir. DeMars (2010, s. 36, 80) madde ayırt edicilik değeri yüksek olan maddelerin θ kestiriminde daha faydalı olacağını ve *a* parametresinin yüksek olduğu durumlarda madde bilgi fonksiyonunun da daha yüksek olacağını belirtmektedir. 4PLM altında θ'ların diğer modellerdekine göre daha az hatayla kestirilmesi bu model altında kestirilen *a* parametre değerlerinin hayli yüksek olması ile de desteklenmektedir. Ayrıca 4PLM için hesaplanan madde bilgi fonksiyonları en üst seviyeye gelmiştir.

3PLM ve 4PLM altında kestirilen *c* parametresine bakıldığında birçok maddenin sıfırdan faklı değerler aldığı, bu maddelerin şansla doğru yanıtlanma olasılıklarının bulunduğu görülmektedir. Lord (1974) iyi geliştirilen standart testlerde iyi çeldiricilerin olduğunu ve düşük yetenek düzeyindeki yanıtlayıcıları çektiğini; bu nedenle *c* parametresinin daha düşük olma eğiliminde olduğunu belirtmektedir.

4PLM altında kestirilen *d* parametresinin 1,00'dan farklı olması, yüksek yetenekli yanıtlayıcıların dikkatsizlik ve benzeri sebeplerle doğru seçenekten başka seçeneğe yönelmeleri olarak ifade edilmektedir. Bu durum, madde kökündeki karmaşık ifadeler ya da anlatım bozukluğundan da kaynaklanabilmektedir. 4PLM altında kestirim yapıldığında birçok madde için *d* parametresine ilişkin kestirimlerin 1,00'dan düşük olduğu görülmüştür. Bu durum, ilgili test maddeleri ile yanıtlayıcılara ilişkin yetenek kestirimi yapıldığında *d* parametresini dikkate alan 4PLM altında kestirim yapılması gerekliliğini ortaya koymaktadır. Ayrıca *c* ve *d* parametrelerine ilişkin standart hatalar oldukça düşüktür ve bu, parametrelerin doğru bir şekilde kestirildiğini göstermektedir. Madde parametre kestirimleri ve standart hataları dikkate alındığında bu çalışmadaki ilgili veri seti için *c* ve *d* parametre kestirimlerinin yapıldığı modellere ihtiyaç olduğu sonucuna ulaşılmıştır.

Araştırmada ele alınan dört model altında ayrı ayrı yetenek parametreleri ve standart hatalar kestirilmiştir. Standart hatalar arası farkın olup olmadığının test edilmesi için yapılan analizlerden elde edilen bulgulara dayalı olarak yetenek parametresi kestirimlerinin doğruluk düzeylerinin modellere göre farklılaştığı sonucuna varılmıştır. Yapılan ikili karşılaştırma testleri sonucu, yetenek parametresi kestiriminin en az hatalı şekilde 4PLM altında yapıldığı bulunmuştur. 4PLM altında maddelere ilişkin ayırt edicilik parametre kestirimlerinin en yüksek olması ve 4PLM'nin en yüksek bilgiyi sağlaması da yetenek kestiriminin en az hata ile bu model altında kestirilmesini destekler niteliktedir. Benzer şekilde yapılan çalışmalarda 3PLM yerine 4PLM kullanıldığında kestirimin standart hatasının düştüğü ve dolayısıyla kestirimin doğruluğunun arttığı belirtilmektedir (Liao, Ho, ve Yen, 2012; Rulison ve Loken, 2009; Yen, Ho, Liao, ve Chen, 2012; Yen, Ho, Liao, Chen, ve Kuo, 2012).

Test bilgi fonksiyonları testin dört model altında da makul düzeyde bilgi sağlayıcı olduğunu göstermektedir. Test dört model altında da alt-orta (θ = [-1,06] - [0,08]) yetenek düzeyi için en fazla bilgiyi sağlamıştır. Bu çalışmadaki ilgili test maddeleri ile öğrencilerin başarı açıcından seviyelerinin belirlenmesi amaçlandığı için güçlüğü orta düzeyde olan maddelerin çok olması ve bu bağlamda orta yetenek düzeyi için daha fazla bilgi sağlaması beklenmektedir. Orta yetenek düzeyine yakın en çok bilgi veren model 4PLM'dir. Testin 21 maddelik formu en çok bilgiyi de 4PLM model altında sağlamaktadır. Bilgi miktarı ve standart hata arasındaki ilişki dikkate alındığında, en fazla bilgi sağlayan 4PLM'nin en düşük hata oranı ile ilgili modeller arasında en doğru kestirimi yaptığı sonucuna ulaşılmaktadır. Ayrıca bu araştırmanın bulguları ile benzer şekilde Waller ve Reise (2010) 4PLM altında yapılan kestirimin daha fazla bilgi sağladığını belirtmişlerdir.

5. KAYNAKÇA

Baker, F. B. (2001). *The Basics of Item Response Theory*. United States of America: ERIC Clearinghouse on Assessment and Evaluation.

Baykul, Y. (2010). *Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması*. Ankara: Pegem.

Brown, T. (2015). *Confirmatory factor Analysis for applied Research*. New York ve London: Guilford.

Chen, W.-H., & Thissen, D. (1997). Local Dependence Indexes for Item Pairs Using Item Response Theory. *Journal of Educational and Behavioral Statistics*, *22*(3), 265–289. https://doi.org/10.3102/10769986022003265

Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. California: Thomson Learning.

De Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory*. *Methodology in the Social Sciences*. New York: Guildford.

DeMars, C. (2010). *Item Response Theory*. New York: Oxford University Press.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. New Jersey: Lawrence Erlbaum.

Fleck, M. P., Poirier-Littre, M. F., Guelfi, J. D., Bourdel, M. C., & Loo, H. (1995). Factorial structure of the 17-item Hamilton Depression Rating Scale. *Acta Psychiatrica Scandinavica*, *92*(3), 168–172. http://dx.doi.org/10.1111/j.1600-0447.1995.tb09562.x

Hambleton, R. K., & Jones, R. W. (1993). Comparison of Classical Test Theory and Item Response Theory and Their Applications to Test Development. *Educational Measurement: Issues and Practice*, *12*(3), 38–47. http://dx.doi.org/10.1111/j.1745-3992.1993.tb00543.x

Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Boston: Kluwer Nijhoff.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. London: Sage.

Harvey, R. J., & Hammer, A. L. (1999). Item Response theory. *Counseling Psychologist*, *27*(3), 353–374. https://doi.org/10.1177/0011000099273004

Houben, R. M. A., Leeuw, M., Vlaeyen, J. W. S., Goubert, L., & Picavet, H. S. J. (2005). Fear of movement/injury in the general population: Factor structure and psychometric properties of an adapted version of the Tampa Scale for Kinesiophobia. *Journal of Behavioral Medicine*, *28*(5), 415–424. https://doi.org/10.1007/s10865-005-9011-x

Kim, D., De Ayala, R. J., Ferdous, A. A., & Nering, M. L. (2007). *Assessing Relative Performance of Local Item Dependence (LID) Indexes*. Chicago: Paper presented at the annual meeting of the National Council on Measurement in Education.

Liao, W., Ho, R., & Yen, Y. (2012). The Four-Parameter Logistic Item Response Theory Model as a Robust Method of Estimating Ability Despite Aberrant Responses. *Social Behavior and Personality*, *40*(10), 1679–1694. DOI: 10.2224/sbp.2012.40.10.1679

Loken, E., & Rulison, K. L. (2010). Estimation of a four-parameter item response theory model. *The British Journal of Mathematical and Statistical Psychology*, *63*(3), 509–25. https://doi.org/10.1348/000711009X474502

Lord, F. M. (1974). Estimation of latent ability there are omitted responses. *Psychometrika*, *39*(2), 247–

264. http://www.jstor.org/stable/20461894 adresinden erişilmiştir.

Magis, D. (2013). A Note on the Item Information Function of the Four-Parameter Logistic Model. *Applied Psychological Measurement*, *37*(4), 304–315. https://doi.org/10.1177/0146621613475471

Maydeu-Olivares, A., Cai, L., & Hernández, A. (2011). Comparing the Fit of Item Response Theory and Factor Analysis Models. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*(3), 333–356. https://doi.org/10.1080/10705511.2011.581993

Reise, S. P., & Waller, N. G. (2003). How many IRT parameters does it take to model psychopathology items? *Psychological Methods*, *8*(2), 164–184. https://doi.org/10.1037/1082-989X.8.2.164

Romera, I., Delgado-Cohen, H., Perez, T., Caballero, L., & Gilaberte, I. (2008). Factor analysis of the Zung self-rating depression scale in a large sample of patients with major depressive disorder in primary care. *BMC Psychiatry*, *8*, 4. https://doi.org/10.1186/1471-244X-8-4

Rulison, K. L., & Loken, E. (2009). I've Fallen and I Can't Get Up: Can High-Ability Students Recover From Early Mistakes in CAT? *Applied Psychological Measurement*, *33*(2), 83–101. https://doi.org/10.1177/0146621608324023

Rupp, A. A. (2003). Item response modeling with BILOG-MG and MULTILOG for windows. *International Journal of Testing*, *3*(4), 365–384. DOI: 10.1207/S15327574IJT0304_5

Spearman, C. (1904). Spearman's rank correlation coefficient. *The American Journal of Psychology*, *15*, 72–101.

Spearman, C. (1907). Demonstration of Formulæ for True Measurement of Correlation. *The American Journal of Psychology*, *18*(2), 161–169. doi:10.2307/1412408

Turgut, M. F., & Baykul, Y. (1992). Ölçekleme Teknikleri. Ankara: Ösym.

Waller, N. G., & Reise, S. P. (2010). Measuring Psychopathology with Nonstandard Item Response Theory Models : Fitting the Four-Parameter Model to the Minnesota Multiphasic Personality Inventory. S. E. Embretson (Ed.), *Measuring psychological constructs: Advances in model-based approaches* (147-173) içinde. Washington, DC, US: American Psychological Association. http://dx.doi.org/10.1037/12074-007

Yen, W. M. (1984). Effects of Local Item Dependence on the Fit and Equating Performance of the Three-Parameter Logistic Model. *Applied Psychological Measurement*, *8*(2), 125–145. https://doi.org/10.1177/014662168400800201

Yen, Y.-C., Ho, R.-G., Liao, W.-W., Chen, L.-J., & Kuo, C.-C. (2012). An Empirical Evaluation of the Slip Correction in the Four Parameter Logistic Models With Computerized Adaptive Testing. *Applied Psychological Measurement*, *36*(2), 75–87. https://doi.org/10.1177/0146621611432862

Yen, Y., Ho, R., Liao, W., & Chen, L. (2012). Reducing the Impact of Inappropriate Items on Reviewable Computerized Adaptive Testing. *Educational Technology & Society*, *15*, 231–243.

Zenisky, A. L., Hambleton, R. K., & Sireci, S. G. (2002). Effects of Local Item Dependence on the Validity of IRT Item, Test, and Ability Statistics. *Journal of Educational Measurement*, *39*(4), 1–16. https://eric.ed.gov/?id=ED462426 sayfasından erişilmiştir.