Turkish Studies
Volume 14 Issue 1, 2019, p. 181-203
DOI: 10.7827/TurkishStudies.14759
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. Ömay ÇOKLUK BÖKEOĞLU – Doç. Dr. Kaan Zülfikar DENİZ*

This article was checked by iThenticate.

İKİ KATEGORİLİ MADDE TEPKİ KURAMI MODELLERİ İLE SINIFLAMALI TEPKİ MODELİNİN KARŞILAŞTIRILMASI*

*Serpil ÇELİKTEN*** *- Emine ÖNEN****

ÖZET

Bu araştırmanın amacı, ikili kodlama yoluyla puanlanan maddeler için madde tepki kuramı modelleri ile çoklu puanlanan sınıflamalı tepki modelini model-veri uyumu, yetenek kestiriminin doğruluğu ve test bilgi fonksiyonları açısından karşılaştırmalı olarak incelemektir. Çalışma grubu 2012 Seviye Belirleme Sınavına katılan öğrencilerden seçilen 1500 öğrenciden oluşmaktadır. Araştırma amacı doğrultusunda istatistiksel analizler R Studio yazılımı, Multilog, Lisrel 8.80 ve SPSS paket programları kullanılarak yapılmıştır. Model uyumunu test etmeye yönelik olarak yapılan analizler sonucunda genel uyum ve madde uyumu bazında STM, 2PLM ve 3PLM'nin araştırma verisine yeterli düzeyde uyum sağladığı, bu modeller arasında en yüksek uyumu sağlayan modelin ise 3PLM olduğu gözlenmiştir. Yetenek parametrelerinin kestirimi açısından ise STM' nin ikili kodlama yoluyla puanlanan MTK modellerine göre daha doğru yetenek kestirimlerine yol açtığı bulunmuştur. Test bilgi fonksiyonları dikkate alındığında ise STM'nin düşük yetenek ranjında, incelenen bütün ikili kodlama modellerinden daha fazla bilgi sağladığı gözlenmiştir. Ancak her ne kadar STM' ye dayalı olarak yetenek kestirimleri daha doğru yapılmış olsa da, 3PLM' ye göre yetenek kestirimi konusunda daha az bilgi sağladığı gözlenmiştir. Bununla birlikte STM'nin, çeldiricilerin etkililiği de incelenmek koşuluyla, çoktan seçmeli test maddelerinin tüm yönleriyle incelenip geliştirilmesine katkı sağlayacağı düşünülmektedir. Bu doğrultuda STM'nin, özellikle uygulamada madde bankası oluşturma çalışmaları için yararlı olacağı düşünülmektedir.

---

* Bu makale Serpil ÇELİKTEN tarafından Dr. Öğr. Üyesi Emine ÖNEN'in danışmanlığında hazırlanmış "İki Kategorili Madde Tepki Kuramı Modelleri ile Sınıflamalı Tepki Modelinin Karşılaştırılması" isimli Yüksek Lisans tezinden üretilmiştir.

** Arş. Gör., Dicle Üniversitesi, Ziya Gökalp Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, E-posta: sserpilcelikten@gmail.com

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

Anahtar Kelimeler: Çoktan seçmeli test, Madde Tepki Kuramı, İki Kategorili MTK modelleri, Sınıflamalı Tepki Modeli

COMPARISON OF DICHOTOMOUS ITEM RESPONSE THEORY MODELS AND NOMINAL RESPONSE MODEL

ABSTRACT

The purpose of this study is to examine the dichotomous item response theory models and nominal response model in terms of model-data fit, accuracy of ability estimation, and test information functions in a comparative way. The study group was composed of 1500 students taken 2012-SBS. For the research purposes, the analysis were conducted through the R Studio software, Multilog, Lisrel 8.80 and SPSS. In terms of model data fit, it was observed that nominal response model, 2 parameter logistic model and 3 parameter logistic model fits the data both in terms of overall fit and item fit, and 3PLM is the best fitting model among them. The abilities were most precisely estimated with NRM. Lastly, in terms of test information functions, NRM provided the most information than the dichotomus IRT models at the low ability range. However, NRM provided less information in terms of total test information function than 3PLM, although the ability parameters were estimated with better accuracy with NRM than 3PLM. However, it is considered that nominal response model would be beneficial to be improved of items with all aspcets by examining the effectiveness of distractors. Therefore, it is believed that nominal response model, especially in practice, would contribute to studies aimed at constructing item bank.

STRUCTURED ABSTRACT

Introduction

There are various types of item response theory models to model the educational and pschological data. Item response theory models are categorized in many ways according to the use and mathematical backgrounds of these models. One of these categorization is considered according to the type of scoring the items. In this way, item response theory models are called dichotomous and polytomous item response theory models. 1 parameter logistic model (1PLM), 2 parameter logistic model (2PLM) and 3 parameter logistic model (3PLM) are the most common dichotomous models. Dichotomous item response theory models are used to model the data in which only two response categories are available such as correct and incorrect. Especially in the tests composed of multiple choice items, dichotomous item response theory models are used commonly. In these testing situations, correct answer is coded as 1, and all other distractors are coded 0. However, it is a challange whether there are significant information in the answers of the participants given to the distractors. To response this questions, polytomous item response theory models were developed. Nominal response model (NRM) is one of the polytomous item response theory models that could be used to model the distractors and get the partial

information from these distractors so as to provide more accurate ability estimations (Bock, 1972).

Multiple choice items are commonly used in testing situations and crucial decisions are made about the students depending on these tests. Therefore, it is important to conduct the testing prosess as accurate as possible. In this direction, NRM could be an important model to model the multiple choice data and make estimations before the crucial decisions because it takes into account the distractors while making the estimations.

The purpose of this study is to examine the dichotomous item response theory models and nominal response model in terms of model-data fit, accuracy of ability estimation, and test information functions in a comparative way.

Method

Study Group

The study group was composed of 1500 students participated to 2012-SBS. Data was obtained from the Ministry of Education. Data includes the mathematic subtest of SBS. Study was conducted on the item response pattern of this subtest.

Instrument

Study was conducted on the 8th grade students' item response patterns of mathematic subtest of 2012-SBS. Mathematic subtest is composed of 20 multiple choice items with learning fields of numbers, geometry, probability and statistics, measurement and algebra (MEB, 2009).

Data Analysis

To conduct the unidimensional item response theory analysis, assumptions of unidimensionality, local independence and monotonic increases of the curves were checked. After testing the assumptions, model-data fit for each model in terms of both overall fit and item fit were investigated. For the investigation of overall fit, RMSEA, TLI, CFI and SRMRS indices were evaluated and for the investigation item fit, S-X^2 statistics were used. Then ability estimations were obtained under each model and they were investigated in a comparative way. To test whether the means of standart errors differs from each other according to the models, One-way analysis of variance for repeated measures were used. Moreover, the test information functions were obtained as general information through the theta scale between [-3,3] under each model. Besides the general test information functions, by taking three different theta intervals account, three ability levels are considered as low between [-3, -1], middle between (-1, 1) and high between [1, 3]. In this way, amount of test information were calculated at three levels. All these analysis were conducted through SPSS, LISREL, Multilog and R studio with many packages (mcIRT, sirt, mirt, psych R packages).

Results

When all models are considered according to the evaluation of overall fit, it was seen that the order of the models starting with the best fit is as 3PLM, 2PLM, NRM and 1PLM. Moreover, fit indices of the NRM

and 2PLM are close to each other especially for the RMSEA values. Besides the overall fit, item fit indices were investigated under each model by using the S-X^2 statistics. According to the results of item fit, it was seen that all items fit the model under 3PLM and the 80 % of the items fit to data under NRM and 2PLM and these well-fitting items are not identical. However, most of the items present poor fit to the data under 1PLM.

Besides the model-data fit, accuracy of the ability estimations were investigated through use of the standart errors of these estimations. According to the evaluations of the standart errors, it was seen that the abilities were most accurately estimated with NRM. It was followed by the 2PLM, 3PLM, and 1PLM.

Lastly, in terms of test information functions, models are investigated by considering the three different ability levels as low, middle and high besides the general test information functions provided by each model. When models were generally evaluated in terms of the total test information through the scale, it was seen that 3PLM provides the most information although the ability parameters were estimated more accurately with NRM than 3PLM. In terms of total test information functions, 3PLM is followed by NRM, 2PLM and 1PLM. When considering the ability levels, it was seen that NRM provided the most information than the dichotomus IRT models at the low ability level. At the middle ability level, 3PLM provided the most information and at the high level 1PLM provides the most information. Moreover, the amount of information provided by NRM and 2PLM is closest in terms of general information and information provided at each ability level.

Discussion and Conclusion

As a general result of evaluation of model data fit, it could be concluded that NRM, 2PLM and 3PLM fits the data both in terms of overall fit and item fit, and 3PLM is the best fitting model among them. However, NRM provides the more accurate ability estimations although the amount of information through the scale provided by 3PLM is more than NRM. Besides these results, investigations of the test information functions reveal the fact that each model behaves differently when different ability levels are considered. Therefore, it could be crucial to decide the appropriate model according to the distributional features of the data. Moreover, since NRM provides a clearer picture of the items, it is considered that nominal response model would be beneficial to provide improved of items with all aspcets by examining the effectiveness of distractors. Therefore, besides the accuracy of the ability estimations, it is believed that NRM, especially in practice, would contribute to studies aimed at constructing item bank.

Keywords: Multiple choice items, Item Response Theory, Dichotomous IRT Models, Nominal Response Model

Giriş

Ulusal ve uluslararası düzeyde, başarıyı ve yeteneği ölçmek amacıyla uygulanan standart testlerde yaygın olarak çoktan seçmeli maddeler kullanılmaktadır (Osterlind, 1998). Bu testlerden elde edilen sonuçlara dayalı olarak birey hakkında tanılayıcı ve yerleştirmeye yönelik kararlar verilmekte,

çoktan seçmeli testlerden elde edilen puanlara göre bireyler çeşitli eğitim kurumlarına seçilmekte ve yerleştirilmektedir. Bu kararlar bireylerin hayatlarına yön verici nitelikte olduğundan, yüksek doğrulukta olmaları son derece önemlidir. Bu doğrultuda da bu tür maddelerden oluşan testlerin uygun şekilde geliştirilmesi, puanlanması ve yorumlanması önem kazanmaktadır. Çoktan seçmeli maddelere dayalı olarak bireyin ölçülen özelliklerine ilişkin elde edilecek bilgilerin niteliği, maddelerin geliştirilme sürecinin yanısıra puanlama sürecinden de oldukça etkilenmektedir. Bu maddelerden elde edilecek bilgiler ve niteliği, çoktan seçmeli maddelerin puanlanma yöntemine dayalı olarak değişebilmektedir.

Çoktan seçmeli test maddeleri, genel olarak, tepki kategorilerinden bir tanesine doğru ve diğer tepki kategorilerine, bir başka deyişle çeldiricilere, eşit düzeyde yanlışmış gibi işlem yapılarak puanlanmaktadır. Klasik Test Kuramı (KTK) çerçevesinde çoktan seçmeli maddeler için genel olarak, her bir doğru yanıta "1" puan verilmekte ve test puanı, bu puanların toplamından elde edilmektedir (DeMars, 2008). Madde Tepki Kuramında (MTK) ise çoktan seçmeli maddelerden oluşan bir test, tek boyutlu ikili puanlanan maddeler için MTK modelleri kullanılarak puanlanabilmektedir (Birnbaum, 1968). Bu MTK modellerinde, çeldiriciler tek bir kategori gibi düşünülmekte ve doğru tepki olasılığı, gizil yeteneğin bir fonksiyonu olarak tanımlanmaktadır. Bu yolla, tek boyutlu ikili puanlanan maddeler için MTK modellerinde madde tepkisine iki düzeyli kategorik bir değişken gibi işlem yapılmaktadır (Penfield & de la Torre, 2008).

MTK'da model testinde maddeleri ve testleri nitelemek, test maddelerini seçmek ve testleri karşılaştırmak amacıyla model ve yetenek parametrelerinin kestirimlerinin yanısıra testin bütününün, bireylerin yetenek düzeylerini (θ) kestirmeye katkısına işaret eden test bilgi fonksiyonu hesaplanmaktadır. Birnbaum (1968) tarafından geliştirilen test bilgi fonksiyonu, aşağıdaki Eşitlik (1) yoluyla elde edilmektedir:

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

Test bilgi fonksiyonu, yetenek parametresinin kestiriminin doğruluğuna ilişkindir (Hambleton, Swaminathan & Rogers, 1991, s. 91, 94).

Madde Tepki Kuramında modeller, ölçülen özelliğin boyut sayısına göre tek boyutlu ve çok boyutlu MTK modelleri olarak iki kategoride ele alınmaktadır. Tek boyutlu MTK modelleri ise, modelde kestirilen parametre sayısına bağlı olarak nitelendirilmektedir. Sadece madde ayırt edicilik parametresinin (a) kestirildiği 1 Parametreli Lojistik Model (1PLM), madde ayırt edicilik (a) ve madde güçlük (b) parametrelerinin kestirildiği 2 Parametreli Lojistik Model (2PLM) ve madde ayırt ediciliği (a), madde güçlüğü (b) ve şans (c) parametrelerinin kestirildiği 3 Parametreli Lojistik Model (3PLM) en yaygın kullanılan modellerdendir.

Bir Parametreli Lojistik Modelde madde ayırt edicilik parametresinin bütün maddeler için aynı olduğu kabul edildiğinden, modelde sadece madde güçlük parametresi kestirilmektedir. Modelin matematiksel gösterimi Eşitlik (2)'de görülmektedir:

$$P_i(\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}} \qquad (2)$$

Belirli bir yetenek düzeyinde (θ) yer alan ve seçkisiz olarak seçilen bir yanıtlayıcının i maddesine doğru yanıt verme olasılığı $P_i(\theta)$ ile temsil edilmektedir. Eşitlikte yer alan b_i, i maddesine ilişkin madde güçlük parametresidir. Parametrenin sayısal değerinin yükselmesi, maddeye doğru yanıt vermek için gerekli yetenek düzeyinin de o kadar yüksek olacağı anlamına gelmektedir. (Hambleton, ve diğ. 1991, s.13). Modelde yer alan "*e"* terimi ise sabit bir değerdir ve *e*=2.718 olarak ele alınır.

İlk olarak Lord (1952) tarafından geliştirilen İki Parametreli Lojistik Model (2PLM) ise yığmalı normal dağılım yaklaşımına dayanmaktadır (Akt: Lord & Novick, 1968). Ancak sonrasında Birnbaum (1968) yığmalı dağılım yaklaşımı yerine madde karakteristik fonksiyonunu kullanarak modeli yeniden tanımlamıştır. Modelin matematiksel gösterimi Eşitlik (3)'de görülmektedir:

$$P_i(\theta) = \frac{e^{D_{ai}(\theta - b_i)}}{1+e^{D_{ai}(\theta - b_i)}} \quad (3)$$

Bu modelde de b_i madde güçlük parametresini nitelendirirken, 1PLM' den farklı olarak burada modele madde ayırt edicilik parametresi (a_i) eklenmiştir. Maddelerin eğim değerinin yüksek olması, maddenin eğimi düşük olan maddelere göre yetenek düzeyleri arasında daha iyi ayırım yaptığını göstermektedir (Hambleton ve diğ., 1991, s. 15-16). Bu modelde lojistik fonksiyonları, normal ogive fonksiyonuna yaklaştırmak amacıyla D sabiti (D=1.7) ya da ölçek sabiti olarak da adlandırılan sabit bir değerden yararlanılmaktadır (Hambleton ve diğ., 1991, s.15).

Birnbaum (1968), 2PLM'e şans parametresini dâhil edip modeli yeniden düzenleyerek; maddeye şansa dayalı olarak yanıt verme durumunun, maddeyi doğru yanıtlama olasılığına katkısının incelenmesini sağlayan 3PLM'i geliştirmiştir. Modele ilişkin matematiksel gösterim Eşitlik (4)'de görülmektedir:

$$P_i(\theta) = c_i + (1 - c_i)\frac{e^{D_{ai}(\theta - b_i)}}{1+e^{D_{ai}(\theta - b_i)}} \quad (4)$$

Eşitlikte c_i (pseudo-chance-level parameter) ile gösterilen şans parametresi, düşük yetenek düzeyindeki bireylerin maddeyi doğru yanıtlama olasılığını temsil etmekte ve düşük asimptotu sıfırdan farklı olan madde karakteristik eğrileri sağlamaktadır (Hambleton ve diğ., 1991, s.17). Şans parametresinin yüksek olması durumunda, maddenin ayırt edicilik parametresinin düşme eğiliminde olduğu belirtilmektedir. Bu nedenle bir maddeyi şansla doğru yanıtlama olasılığının yüksek olması, maddenin sağladığı bilgi miktarının da düşmesine yol açmaktadır (Harvey & Hammer, 1999).

Çoktan seçmeli test maddeleri için madde tepkilerini modellemede tek boyutlu ikili puanlanan maddeler için MTK modellerinden 1PLM, 2PLM ve 3PLM yaygın bir şekilde kullanılmasına rağmen, hem bu modeller hem de KTK'da uygulanan yöntem, doğru yanıtı seçmede başarısız olan bir yanıtlayıcının hangi çeldiriciyi seçtiğini dikkate almamaktadır (DeMars, 2008). Her iki kuram çerçevesinde de çoktan seçmeli maddelerin ikili puanlanması sonucu bireyin doğru yanıtı işaretlememesi, bireyin o maddeye ilişkin bilgisini yok sayarak, ilgili maddeden "0" puan almasına yol açmaktadır. Ancak birey bir maddeye yanlış yanıt verdiğinde bile, seçtiği çeldiriciye dayalı olarak, öğrencinin yeteneği hakkında bilgi edinilmesi mümkündür. Çoktan seçmeli test maddelerindeki çeldiriciler, belirli problem çözme stratejilerinin kullanımına, kavram yanılgılarına ya da hatalı işlemsel kuralları uygulamaya işaret edecek şekilde hazırlanabilir. Dolayısıyla çeldiriciler de aslında doğru yanıtlar kadar, bireyin yeteneği hakkında bilgi içeriyor olabilir (DeMars, 2008). Mislevy (1995), eğitim ortamlarında kullanılan birçok test için her bir maddede yanıtlayıcıların seçtikleri belirli tepki kategorilerinin, toplam test puanlarında ya da MTK dayanaklı yetenek kestirimlerinde açıkça görülmeyen (belli olmayan) bazı bilgiler sağladığını belirtmektedir.

Madde Tepki Kuramı çerçevesinde çoktan seçmeli madde tepkilerini modellemek için kullanılan tek boyutlu çoklu puanlanan maddeler için MTK modellerinden Sınıflamalı Tepki Modeli (Bock, 1972) ve Samejima (1979) ve Thissen ve Steinberg' ün (1984) çoktan seçmeli modelleri ise, doğru yanıttan olduğu kadar çeldiricilerden de bilgi elde etme olanağı sağlamaktadır (akt: Thissen & Steinberg, 1986). Tek boyutlu çoklu puanlanan maddeler için MTK modelleri, tüm tepki seçenekleri arasındaki farkı korumakta ve gizil yeteneğin bir fonksiyonu olarak her bir tepki kategorisini seçme

olasılığını tanımlamaktadır. Bu modeller, tepki seçeneklerinin belirli yetenek düzeyleri için göreli çekiciliği gibi tek tek test maddelerinin işleyişine ilişkin daha derin bir bilgi sağlamakla birlikte (Thissen, Steinberg ve Fitzpatrick, 1989), yanıtlayıcı yeteneğinin daha kesin (precise) kestirimleri ile sonuçlanabilmektedir (Baker, 1992). Bu çalışmada, tek boyutlu çoklu puanlanan maddeler için MTK modellerinden biri olan Sınıflamalı Tepki Modeli (Nominal Response Model) ile 1PLM, 2PLM ve 3PLM; test bilgi fonksiyonları; yetenek parametrelerinin kestirim doğruluğu ve model-veri uyumu bağlamında karşılaştırılmıştır.

Sınıflamalı Tepki Modeli (STM), Bock (1972) tarafından, iki ya da daha fazla sınıflamanın söz konusu olduğu çok kategorili maddelere verilen tepkileri modellemek amacıyla geliştirilmiştir. Çoktan seçmeli test maddelerinde yanıt kategorileri, birbirini karşılıklı olarak dışlayan (mutually exclusive) ve farklı içeriklere sahip kategorilerdir. STM ise, bu yanıt kategorilerinin her birine tepki verme olasılığını bireyin yeteneğiyle olan ilişkisi bağlamında inceleme olanağı sağlamaktadır (DeAyala, 2009; Ostini & Nering, 2006, s.16-18). STM, her bir çeldiriciye özgü bilgiyi birleştirmek yoluyla, gizil yeteneğe ilişkin maksimum bilgiye ulaşmayı sağlayabilmektedir. Bu doğrultuda STM, özellikle de gizil yetenek ölçeğinin (latent trait continuum) en düşük noktasında, ikili puanlanan maddeler için MTK modellerine kıyasla, gizil yeteneğin daha kesin (precise) bir kestiriminin elde edilmesine yol açabilmektedir (Bock, 1972; De Ayala, 1989, 1992; Thissen & Steinberg, 1986). Bu modelde her bir yanıt kategorisine verilen tepkileri bireysel olarak modellemek amaçlandığından, model parametreleri madde bazında değil, maddelerin her bir kategorisi bazında ele alınmaktadır. Dolayısıyla modelin, Eşitlik (5) ve Eşitlik (6)'da sunulan matematiksel gösterimi incelendiğinde de anlaşılacağı gibi, modelde kestirilecek parametre sayısı, yanıt kategori sayısının iki katıdır:

$$P_{ih} = \frac{e^{Zih}}{\sum_{h}^{m} e^{Zih}} \qquad h= 1, 2, \ldots.g\ldots..m. \qquad (5)$$

$$Zih = \zeta + \lambda\theta \qquad (6)$$

Eşitlik (6) incelendiğinde, STM' de, temelde ζ (zeta) ve λ (lambda) olmak üzere iki parametre kestirildiği anlaşılmaktadır. Zeta (ζ) parametresi, sabit parametresidir (intercept parameter) ve *i* maddesinin h. yanıt kategorisinin genel popularitesini göstermektedir. Bir takım doğrusal dönüşümler uygulanması yoluyla bu parametre, iyi bilinen b parametresine dönüştürülmektedir. Baker (1992) bu dönüşüm için Eşitlik (7)'nin kullanılmasını önermektedir:

$$b = -\frac{\lambda}{\zeta} \qquad (7)$$

Alanyazında (Baker, 1992; DeAyala, 1993) bu parametrenin doğrudan güçlük parametresi olarak yorumlanmadığı, bunun yerine yer parametresi olarak ele alındığı vurgulanmaktadır. Bununla birlikte DeAyala (1993), bu yer parametresini, ardışık kategoriler için elde edilen madde kategori yanıt fonksiyonunun yetenek ölçeği üzerindeki yerine göre tanımlamaktadır. Bu tanımlama doğrultusunda Zeta parametresine Eşitlik (7)'de yer alan dönüşümün uygulanması yoluyla b parametresi elde edilmektedir:

$$b = \frac{\zeta(h-1) - \zeta(h)}{\lambda(h) - \lambda(h-1)} \qquad (8)$$

Bu eşitlikler ile doğrusal dönüşümler elde edildiği için Zeta ve b parametrelerine ilişkin değerler paralellik göstermekte, ancak yukarıda da vurgulandığı gibi, b parametresi doğrudan güçlük parametresi olarak yorumlanmayıp, yer parametresi olarak ele alınmaktadır. "λ" parametresi ise, yeteneği çoklu logit ile ilişkilendiren eğim olarak nitelendirilmektedir ve madde ayırt ediciliğini temsil etmektedir. Bu parametreye uygulanan dönüşüm ise Eşitlik (9)'da yer almaktadır:

$$\lambda = a \qquad (9)$$

Eşitlik (9)'dan da anlaşılacağı gibi λ parametresi doğrudan ayırt edicilik olarak ele alınıp yorumlanmaktadır (Bock, 1997, s.33-37; Ostini & Nering, 2006, s.17-18).

Sınıflamalı Tepki Modeli, bireyin bilgisinin yalnızca varlığını ve yokluğunu dikkate alan iki kategorili modeller olarak ele alınan 1PLM, 2PLM ve 3PLM' den farklı olarak, bireyin kısmi bilgisini de dikkate almaktadır (Bock, 1972). İki kategorili modeller ile yanıtlayıcının yeteneği ile her bir kategoriye tepki verme arasındaki ilişki incelenememektedir. Örneğin kısmi olarak doğru yanıtlanmış bir maddeye puan vermek için, ikiden fazla kategori içeren bir modele ihtiyaç bulunmaktadır. Bu noktada STM, ikili puanlanan maddeler için MTK modelleriyle elde edilemeyen kısmi bilgiyi, çoklu puanlama yaparak elde etmeyi sağlamaktadır (DeAyala, 1993; Tomkowicz, 2000). Bu doğrultuda çoktan seçmeli test maddelerine verilen tepkileri modellemede STM'nin kullanılması önerilmektedir. STM ile yanıt örüntüsünün içerdiği bütün bilgiler kullanılabilmektedir ve bu, yetenek kestiriminin doğruluğunun artmasına yol açmaktadır (Tomkowicz, 2000). Bir maddeye ilişkin çeldiricilerin yetenek dağılımı boyunca her bir noktada, bir başka deyişle her bir yetenek düzeyinde seçilebilme olasılığına ilişkin bilgi sağlar. Bununla birlikte alanyazında (DeAyala, 2009, s.238; Stone, 2006), özellikle çeldiriciler öğrencilerin kavram yanılgılarına yönelik olarak hazırlandığında, STM yoluyla bu kavram yanılgıları daha detaylı bir şekilde incelenebileceği belirtilmektedir. Öğrencilerin kavram yanılgılarına yönelik daha detaylı bilgiler elde edilmesi, bunların üstesinden gelmek için yeni öğretim teknikleri geliştirilmesine de yardımcı olabilir. STM daha nitelikli maddeler geliştirilmesine hizmet edebilir. Bu durumun ülkemizde gerçekleştirilmesi amaçlanan madde havuzu oluşturma çalışmalarına önemli katkılarda bulunması beklenmektedir.

Tek boyutlu ikili puanlanan maddeler için MTK modelleri ile STM'nin eğitimsel ve psikolojik testlerden elde edilen verilere uygulanmasına yönelik yapılan çalışmalar (DeAyala, 1989; DeMars, 2008; Şahin, 2012; Tay, Huang & Vermunt, 2016; Tomkowicz, 2000) incelendiğinde, genel olarak MTK kapsamında model-veri uyumu, model parametrelerinin kestiriminin doğruluğu, madde ve test bilgi fonksiyonları, madde ayırımsal işlerliği ve madde yanlılığı konularına odaklanıldığı görülmektedir. Türkiye'de ikili puanlanan maddeler için MTK modellerine ilişkin çok sayıda çalışma (Kılıç, 1999; Çalışkan, 2000; Çelik, 2001; Özkurt, 2002) bulunmasına rağmen, STM'nin incelendiği tek bir çalışmaya (Demirtaşlı, Yalçın & Ayan, 2014) rastlanmıştır. Ancak bu çalışmada da STM, çoktan seçmeli madde tepkilerini modellemek için değil, bir tutum ölçeğinde yer alan maddelere verilen tepkileri modellemek üzere kullanılmıştır. Yurtdışında yapılan çalışmalara bakıldığında STM'nin genellikle simülatif verilerle incelendiği (DeAyala, 1989, 1992; DeAyala & Koch, 1987; DeMars, 2008), gerçek veriye dayalı çalışmaların (DeAyala & Koch, 1988; Drasgow, Levine, Tsien, Williams & Mead., 1995; Tomkowicz, 2000) çok daha sınırlı sayıda olduğu belirlenmiştir.

Sınıflamalı Tepki Modelinin, BUT (CAT) uygulamalarına dayalı diğer MTK modelleriyle karşılaştırıldığı çalışmalarda; STM' ye dayalı yetenek parametre kestirimleri ile 3PLM'ye dayalı yetenek parametre kestirimlerinin yüksek düzeyde ilişkili olduğu, ancak STM altında yetenek parametrelerinin daha düşük standart hata ile kestirildiği (DeAyala ve Koch, 1987), yetenek ölçeği boyunca yetenek kestirimlerinin yanlılık içermediği (DeAyala & Koch, 1988), düşük miktarda bilgi sağladığı yetenek düzeylerinde bile düşük yanlılıkla kestirim yaptığı ve yetenek düzeyi yüksek bireyler için oldukça kesin/net (precise) yetenek kestirimlerinin elde edildiği (DeAyala, 1989) ve STM'nin 3PLM'e kıyasla düşük yetenek düzeyinde ($\theta \leq -1.0$) daha fazla bilgi sağladığı (DeAyala, 1989) sonuçlarına ulaşılmıştır.

Başarı testlerinden elde edilen veriler üzerinde gerçekleştirilen bir çalışmada (Drasgow ve diğ., 1995); Bock' un (1972) sınıflamalı tepki modeli, Samejima'nın (1993) çoktan seçmeli C modeli, Thissen ve Steinberg'in (1984) çoktan seçmeli modeli ve Levine'in (1993) maksimum olabilirlik formül puanlama (maximum-likelihood formula scoring model) modellerinin veriye uyum düzeyleri karşılaştırmıştır. Araştırma bulgularına göre en iyi model-veri uyumu STM için sağlanmıştır. Tomkowicz (2000) ise 1PLM, 2PLM, 3PLM ve STM'yi, düşük-orta ve yüksek olmak üzere üç farklı

yetenek düzeyi için, yetenek parametrelerinin kestirimlerinin doğruluğu açısından karşılaştırmıştır. Araştırma bulguları, lojistik modeller ile STM'nin düşük yetenek düzeyinde yetenek kestiriminin doğruluğu açısından farklılaştığını, STM'nin daha doğru kestirimlerde bulunduğunu açığa çıkarmıştır. Orta ve yüksek yetenek düzeyindeki bireylerin yetenek kestirimlerinin doğruluğu açısından ise modeller birbirine yakın sonuçlar vermektedir. De Mars (2008), 3PLM ve STM'yi karşılaştırdığı çalışmasında 3PLM'nin uygulandığı durumda, düşük yetenek düzeyindeki bireyler için yetenek parametrelerinin (θ) daha yanlı bir şekilde kestirildiğini ve 3PLM altında bu bireyler için test puanlarının güvenirliğinin düşük olduğunu gözlemlemiştir. Çoklu puanlama durumunda ise daha az yanlı yetenek kestirimlerinin yapıldığı belirtilmiştir.

Sınıflamalı Tepki Modelinin, tek boyutlu ikili puanlanan maddeler için MTK modelleriyle karşılaştırıldığı bu araştırmaların bulguları ve STM 'ye yönelik kuramsal açıklamalar doğrultusunda, çoktan seçmeli test maddelerinin, ikili puanlama yapılırken öğrencinin kısmi bilgisini dikkate almadığı için bilgi kaybına neden olduğu sonucuna varılmaktadır. Bu nedenle bu çalışmada, bireyin ölçülen yeteneği hakkında daha fazla bilgiyi ortaya çıkararak daha sağlıklı kararlar alınmasına yol açacağı düşünülen STM ile lojistik modeller; model-veri uyumu, madde parametrelerinin kestirimi ve farklı yetenek düzeylerindeki bireyler için testin sağladığı bilgi miktarı açısından karşılaştırılmıştır. Bu karşılaştırmanın da, testin kullanım amacına hizmet etme derecesini değerlendirme bağlamında, geçerliğe ilişkin kanıt sağlama açısından da önemli olduğu düşünülmektedir.

Bu çalışma öğrenciler hakkında önemli kararların verilmesinde yararlanılan bir başarı testinden elde edilen gerçek veriye dayalı olarak gerçekleştirildiği için, bu çalışmanın uygulamaya da bilgi sağlaması beklenildiğinden ilgili alan yazına katkısı olacağı düşünülmektedir. Araştırma bulgularının aynı zamanda, ülkemizde de sıklıkla kullanılan çoktan seçmeli test maddelerine dayalı olarak öğrencilerin yetenek düzeyine ilişkin daha doğru kestirimler yapılmasına katkıda bulunacağı düşünülmektedir. Dünyada bu testlerden elde edilen sonuçlara dayalı olarak birey hakkında onların yaşamlarını etkileyecek ve yön verecek nitelikte kararlar verildiğinden, birey hakkında doğru kararların verilmesi açısından bireyin yeteneğinin doğru bir şekilde kestirilmesi önem taşımaktadır. Bu durum ölçme ve değerlendirme sürecinin hatalardan olabildiğince arındırılarak yürütülmesi ile gerçekleştirilebilir. Dolayısıyla bu çalışma ölçme ve değerlendirmenin en az hatayla yapılmasına katkıda bulunmak amacıyla farklı yöntemleri karşılaştırmalı olarak değerlendirip en uygun yönteme karar vermeyi amaçladığından önem taşımaktadır. Ayrıca bu çalışmada model-veri uyumu, bilgi fonksiyonları, parametrelerin kestirimlerinin netliği gibi MTK'ya ilişkin birtakım özellikler, farklı modeller altında eş zamanlı olarak karşılaştırılmıştır. Dolayısıyla bu çalışma söz konusu modellere ilişkin daha net bir resmin ortaya çıkarması açısından önem taşımaktadır. Bu doğrultuda bu çalışmanın eğitimsel ve psikolojik ölçmelerde farklı durumlarda uygun modele karar verilmesi açısından araştırmacılara ve birtakım sınavlar doğrultusunda öğrenci hakkında karar vermeye çalışan kurumlara katkı sağlayacağı düşünülmektedir. Tüm bu açıklamaların ışığında bu araştırmanın amacını; Seviye Belirleme Sınavı (SBS)-2012 Matematik alt testinden elde edilen verilere dayalı olarak 1PLM, 2PLM, 3PLM ile STM'nin; model-veri uyumu, yetenek parametre kestirimlerinin doğruluğu, test bilgi fonksiyonlarının sağladığı bilgi miktarı açısından karşılaştırılması oluşturmaktadır.

Yöntem

Araştırma Modeli

Bu çalışmada, 2012 SBS Matematik alt testine ait veri üzerinde 1PLM, 2PLM, 3PLM ve Sınıflamalı Tepki Modeli'ne dayalı olarak model-veri uyumu, birey parametrelerinin kestirimlerinin doğruluğu, test bilgi fonksiyonları karşılaştırmalı olarak incelenmiştir. Bu çalışmada mevcut kuramsal bilgiye yenilerinin eklenmesi ve söz konusu bilginin gelişmesine katkıda bulunulması amaçlandığından temel araştırma niteliğindedir (Fraenkal, Wallen & Hyun, 2012, s.7).

Çalışma Grubu

Çalışma grubunu 2012 yılında uygulanan SBS'ye katılmış öğrencilerden; MEB'den elde edilen 50000 kişilik gruptan seçkisiz olarak seçilen, 1500 öğrenci oluşturmaktadır. Çalışmada ilgili MTK modellerinin model-veri uyumunu madde düzeyinde değerlendirmek için uyum istatistikleri hesaplanmıştır. Çalışma grubunun büyüklüğünü belirlemede, bu çalışmada incelenen MTK modelleri için gerekli kişi sayısı göz önünde bulundurulmuştur. Bu çalışmada incelenen MTK modelleri için örneklemin ya da çalışma grubunun hangi büyüklükte olması gerektiğine ilişkin çalışmalara (de la Torre & Hong, 2010; Foley, 2010) bakıldığında, 1000 kişilik örneklemin/çalışma grubunun yeterli olduğu anlaşılmaktadır. Bu doğrultuda çalışma, MEB'den elde edilen 50000 kişilik veri setinden seçkisiz olarak seçilen 1500 öğrenciden elde edilen veriler üzerinde gerçekleştirilmiştir.

Veri Toplama Araçları

Bu çalışma, SBS-2012 kapsamında 8. Sınıf öğrencilerine uygulanan matematik alt testinde yer alan 20 maddeye ilişkin öğrenci verileri üzerinde gerçekleştirilmiştir. Matematik alt testinde öğrenme alanları; sayılar, geometri, olasılık ve istatistik, ölçme ve cebir şeklindedir (MEB, 2009). Çalışma için gerekli veriler, Milli Eğitim Bakanlığı Yenilik ve Eğitim Teknolojileri tarafından sağlanmıştır. Çalışmada hem ikili puanlanan maddeler için MTK modellerine hem de çoklu puanlanan MTK modellerinden STM'ye dayalı incelemeler yapılmıştır. Bu doğrultuda ham veriler; ikili puanlanan maddeler için MTK modelleri için doğru yanıt 1, diğer bütün yanıtlar 0 olacak şekilde 1-0 formatında düzenlenirken, STM için ise uzman görüşleri doğrultusunda çeldiricilerin doğru yanıta olan yakınlık derecesi belirlenmiştir. STM için matematik eğitimi alanında çalışan beş uzmandan alınan görüşler değerlendirilmiş, ortak görüşler doğrultusunda her bir madde için çeldiricilerin doğru yanıta yakınlık-uzaklık dereceleri belirlenmiştir. Bu doğrultuda doğru yanıt 4, doğru yanıta en yakın çeldirici 3, daha az yakın çeldirici 2 ve en uzak çeldirici 1 olacak şekilde kodlamalar yapılarak veri seti düzenlenmiştir.

Bu çalışma MTK çerçevesinde gerçekleştirildiğinden, söz konusu ölçümlerin MTK bağlamında güvenirliğine ilişkin kanıtlar elde etmek amacıyla ölçümlerin marjinal güvenirlik katsayısı hesaplanmıştır. Bu katsayı, gerçek puan modeline dayandırılmakta (Lord & Novick, 1968) ve koşullu standart hata ortalamasına dayalı olarak bir testin genel güvenirliğinin kestirimi olarak ifade edilmektedir. Marjinal güvenirlik katsayısının teorik alt yapısı ile klasik güvenirlik katsayısının teorik alt yapısı karşılaştırılabilir olduğundan, elde edilen katsayılar da benzer şekilde yorumlanabilmektedir (Dimitrov, 2003). 1PLM, 2PLM, 3PLM ve STM altında söz konusu ölçümlere yönelik marjinal güvenirlik katsayıları sırasıyla 0.800, 0.892, 0.856 ve 0.883 olarak hesaplanmıştır. Bu doğrultuda testin 1PLM, 2PLM, 3PLM ve STM altında güvenilir ölçümler sağladığı sonucuna varılmıştır.

Çalışmada ilgili ölçümlerin geçerliğine ilişkin kanıtlar, Açımlayıcı (AFA) ve doğrulayıcı faktör analizlerinin (DFA) uygulanması ile elde edilmiştir. AFA, verilerin 1-0 şeklinde puanlanmasından dolayı, tetrakorik korelasyon matrisine dayalı olarak hesaplanmıştır. AFA, rotasyonsuz "minimum residual" (unweighted least squares) faktör çıkartma tekniği kullanılarak maksimum olabilirlik yöntemi (maximum likelihood) ile gerçekleştirilmiştir. Analiz sonucunda özdeğeri 1'den büyük olan (özdeğer=10.53) ve varyansın %53'ünü açıklayan tek bir faktörün ortaya çıktığı gözlenmiştir. Maddelerin faktör yük değerleri ise λ=0.35 ile 0.86 arasında değişim göstermiştir. AFA'dan elde edilen bulgulara dayalı olarak, matematik alt testinin "Matematik Başarısı" olarak adlandırılabilecek tek faktörlü bir yapıyı ölçtüğü sonucuna ulaşılmıştır. Ardından, DFA yoluyla ölçümlerin geçerliğine ilişkin kanıt elde etmek üzere ilgili ölçümlerin faktör yapısına ilişkin tek faktörlü bir ölçme modeli tanımlanmıştır. Maddelerden elde edilen puanlar 1-0 şeklinde olduğu için madde tepkileri "ordinal" birer değişken olarak tanımlanıp model parametrelerinin kestirimi için "Robust Maksimum Likelihood" yöntemi kullanılmıştır. DFA, gizil değişken ölçeğini tanımlamak amacıyla işaretçi gösterge yöntemi kullanılarak 1. madde için faktör yük değeri 1'e eşitlenerek yapılmıştır. Test edilen tek faktörlü ölçme modeline ilişkin uyum indeksleri (CFI = 1.00, NFI=1.00, GFI=0.92, AGFI=0.90, RMSEA=0.024), söz

konusu modelin çalışma grubundan elde edilen veri setine iyi uyum sergilediğine işaret etmektedir (Brown, 2006). Model parametrelerinin standartlaştırılmış kestirimleri dikkate alındığında, göstergelere ilişkin hesaplanan faktör yük değerlerinin $\lambda = 0.31$ ile 0.87 arasında, özgül varyans değerlerinin ise $\varepsilon = 0.24$ ile 0.90 arasında değiştiği belirlenmiştir. Hem AFA hem de DFA sonuçları, SBS-2012 Matematik alt testinde yer alan maddelerin, "Matematik Başarısı" olarak adlandırılabilecek yapının uygun birer temsilcisi olduklarına ve yapıyı kabul edilebilir düzeyde hata içerecek şekilde ölçebildiğine işaret etmektedir. Sonuç olarak bu bulgular ilgili ölçümlerin yapı geçerliğine ilişkin kanıtlar olarak değerlendirilmiştir.

Bunun yanısıra alt, orta ve yüksek yetenek düzeyi için ayrı ayrı hesaplanan test bilgi fonksiyonları da, her bir model altında, testten sağlanan bilgi miktarının en yüksek olduğu yetenek ranjının orta yetenek düzeyi olduğuna, dolayısıyla ölçümlerin amacına hizmet ettiğine ve bu yetenek ranjında en az hatayla güvenilir ölçümler elde edilebileceğine işaret etmektedir. SBS, seviye belirlemeye yönelik olduğundan, Matematik alt testinin kullanım amacına hizmet ettiği ifade edilebilir.

Verilerin Analizi

Araştırma amacı doğrultusunda model testi öncesinde MTK'nın temel varsayımlarının, çalışma grubundan elde edilen veri seti için karşılanıp karşılanmadığı test edilmiştir. Bu doğrultuda öncelikle tek boyutluluk varsayımının karşılanıp karşılanmadığını test etmek üzere AFA ve DFA gerçekleştirilmiştir. AFA, R Studio yazılım ortamında psych paketi kullanılarak, DFA ise LISREL 8.7 programı kullanılarak yapılmıştır. AFA sonucunda özdeğeri 1'den büyük tek bir faktörün ortaya çıktığının gözlenmesi, DFA ile test edilen tek faktörlü ölçme modelinin çalışma grubundan elde edilen veri setine iyi düzeyde uyum sergilemesi ve hem AFA hem de DFA sonucunda ilgili maddelerin tek faktöre kabul edilebilir düzeyde yüklenmiş olması, veri setinin "tek boyutluluk" varsayımın karşılandığına işaret etmektedir. AFA ve DFA sonuçları, geçerlik kanıtları olarak "Veri Toplama Araçları" bölümünde sunulmuştur.

Madde Tepki Kuramı'nın bir diğer önemli varsayımı "yerel bağımsızlık"tır (Hambleton, Swaminathan & Rogers, 1991). Genel olarak alanyazında (Embretson & Reise, 2000; Hambleton ve diğ., 1991) tek boyutluluk varsayımı karşılandığı durumda "yerel bağımsızlık" varsayımının da karşılanmış olduğu kabul edilmektedir. Bununla birlikte, ilgili madde çiftleri arasındaki bağımlılığın derecesini gösteren bir indeks olan Yen'in (1984) Q3 istatistiği kullanılarak da bu varsayımın test edilmesi mümkün olmaktadır. R Studio programında, sirt R paketi kullanılarak, her bir MTK modeli altında, testte yer alan her bir madde çifti için Q3 istatistiği ayrı ayrı hesaplanmıştır. Madde çiftleri için yerel bağımsızlık varsayımının karşılanması için, bu istatistiğin mutlak değerinin 0.2'den düşük olması gerekmektedir (DeAyala, 2009). Bu doğrultuda yerel bağımsızlık varsayımı, 1PLM, 2PLM ve 3PLM altında 190 madde çifti için test edilmiştir. Değerlerin 1PLM için Q3=-0.149 ile 0.080, 2PLM için Q3=-0156 ile 0.093 ve 3PLM için Q3=-0.152 ile 0.078 arasında değişim göstermesi, üç model altında her bir madde çifti için söz konusu varsayımın karşılandığı şeklinde değerlendirilmiştir.

Madde karakteristik eğrisinin monotonik artiş göstermesi varsayımını test etmek üzere Multilog programından yararlanılmış ve araştırmada test edilen her bir model altında ve her bir madde için madde karakteristik eğrisi elde edilmiştir. Elde edilen madde karakteristik eğrilerinin incelenmesine dayalı olarak, her bir madde için gizil değişken ile madde yanıtları arasında doğrusal bir ilişki olduğu, bir diğer deyişle, yetenek düzeyi arttıkça maddeyi doğru yanıtlama olasılığının da arttığı sonucuna ulaşılmıştır. MTK'nın geçerli sonuçlar vermesi için, hem madde parametrelerinin hem de birey parametrelerinin değişmezliğinin sağlanması gerekmektedir. Madde parametrelerinin değişmezliğini incelemek üzere çalışma grubu seçkisiz olarak eşit iki alt gruba ayrılmıştır. Bu iki gruptan elde edilen veri setleri için ayrı ayrı her bir MTK modeli altında madde parametreleri kestirilerek, iki gruba ilişkin parametre kestirimleri arasındaki ilişkiler, SPSS 20.0 programında, Pearson Momentler Çarpımı Korelasyon tekniği ve saçılma grafikleri aracılığıyla incelenmiştir. İki kategorili MTK modelleri (1PLM, 2PLM ve 3PLM) altında kestirilen b parametresi için hesaplanan korelasyon katsayıları (1PML için r=.979, 2PLM

için r=.995, 3PLM için r=.992) ile 2PLM ve 3PLM altında kestirilen a parametresi için hesaplanan korelasyon katsayılarına (2PLM için r=.996 ve 3PLM için r=.876) ve 3PLM altında kestirilen c parametresi için hesaplanan korelasyon katsayısına (r=.753) dayalı olarak her bir model için madde parametrelerinin değişmezliğinin sağlandığı sonucuna ulaşılmaktadır. Ayrıca saçılma grafikleri incelenmesine dayalı olarak da, her bir model altında kestirilen madde parametreleri arasındaki ilişkilerin doğrusal olduğu sonucuna ulaşılmıştır. STM için de, benzer şekilde işlemler ve analizler yapılmıştır. Bu doğrultuda STM altında her bir maddedeki yanıt kategorileri için ayrı ayrı kestirilen a_1, a_2, a_3 ve a_4 parametrelerinin her birine ilişkin hesaplanan korelasyon katsayılarına (r=.712, r=.550, r=.853, r=.960) dayalı olarak bu parametreler açısından değişmezlik varsayımının sağlandığı sonucuna ulaşılmıştır. Benzer şekilde maddelerin yanıt kategorilerine ilişkin ayrı ayrı kestirilen b_1, b_2, b_3 ve b_4 parametrelerinin her birine ilişkin hesaplanan korelasyon katsayıları da (r=.992, r=.963, r=.981, r=.981), bu parametreler için değişmezlik varsayımın karşılandığına işaret etmektedir (Hambleton & Swaminathan, 1985). Ayrıca saçılma grafikleri ile parametrelerin arasındaki ilişkiler incelenmiş ve bu ilişkilerin doğrusal olduğu sonucuna ulaşılmıştır.

Birey parametrelerinin değişmezliği ise, testin tek ve çift numaralı maddelerinden oluşan iki alt testten elde edilen veri setlerine dayalı olarak her bir model altında kestirilen yetenek parametreleri arasındaki ilişkilere yönelik korelasyon katsayılarının hesaplanması yoluyla test edilmiştir. Hesaplanan bu korelasyon katsayıları; 1PLM (r=.813), 2PLM (r=.820), 3PLM (r=.885) ve STM (r=.820) için birey parametreleri açısından değişmezlik varsayımının karşılandığına işaret etmektedir. Ayrıca bu ilişkilere yönelik oluşturulan saçılma grafiklerinin incelenmesi sonucunda da, söz konusu ilişkilerin doğrusal olduğu sonucuna ulaşılmıştır.

Ardından her bir model altında model-veri uyumu, hem modelin veriye genel uyumu olarak (overall fit) hem de madde düzeyinde uyum olarak incelenmiştir. Genel model uyumu, uyum iyiliği indekslerine dayalı olarak değerlendirilmektedir. Alternatif modellerin uyum düzeyleri, -2 loglikelihood değerlerine dayalı olarak tek yönlü varyans analizi ile karşılaştırılmakta ve elde edilen olasılık değerinin anlamlı olması, daha kompleks modelin (daha fazla parametrenin kestirildiği model) veriye daha iyi uyum sergilediğine işaret etmektedir (DeAyala, 2009, s.140). Model-veri uyumuna ilişkin bütün istatistikler, R Studio ortamında mirt R paketi kullanılarak hesaplanmış ve modellerin uyum düzeylerini karşılaştırmak üzere tek yönlü varyans analizi (ANOVA) de bu paketten yararlanılarak gerçekleştirilmiştir. Öncelikle her bir modele ilişkin uyum istatistikleri (CFI, TLI, RMSEA, SRMRS) elde edilmiştir. Modellerin uyum düzeylerinin -2 loglikelihood değerleri temelinde karşılaştırılabilmesi için bu değerlerin aynı veri setine dayalı olarak elde edilmesi gerekmektedir. Bu doğrultuda bu karşılaştırmalar sadece ikili puanlanan MTK modelleri için gerçekleştirilebilmektedir. Her bir model altında madde düzeyinde uyumu değerlendirmek için R Studio yazılım ortamındaki mirt paketiyle S-χ^2 istatistiği hesaplanmıştır. Bu istatistiğin hesaplanabilmesi için gruplama işlemleri yapılmakta ve kesme puanları kullanılmaktadır. S-χ^2 istatistiğine dayalı olarak madde düzeyinde uyumun değerlendirilmesinde kesme puanı test puanı temelinde belirlenmektedir. Bu istatistik ile hem iki kategorili MTK modelleri, hem de çok kategorili MTK modelleri madde uyumu bazında ele alınabilmektedir.

Model uyumun değerlendirilmesinin ardından, R Studio yazılım ortamında ve Multilog programında her bir model altında madde ve yetenek parametreleri kestirilmiştir. İkili puanlanan MTK modelleri için madde parametreleri sirt paketiyle R studio yazılım ortamında, a parametresi (0-2) aralığına, güçlük parametresi ise ((-2)-2) aralığına sınırlandırılarak (pratik anlamlılık için) kestirilmiştir (Hambleton ve diğ., 1991). STM altında zeta ve lambda parametreleri mcIRT paketiyle R studio yazılım ortamında kestirilirken, bu parametrelerin doğrusal dönüşümü sonucu elde edilen a ve b parametreleri ise Multilog programında kestirilmiştir. Tüm modeller için madde parametre kestirimlerinde Marjinal Maksimum Olabilirlik kestirim yöntemi kullanılmıştır. Tüm modeller altında test bilgi fonksiyonlarının tamamı Multilog programında hesaplanmıştır. Yetenek parametreleri ise, bütün modeller için,

Maksimum Olabilirlik yönteminden yararlanılarak, R studio yazılım ortamında mirt paketi kullanılarak kestirilmiştir. Farklı modeller altında yetenek parametrelerinin kestirimlerinin doğruluğunu karşılaştırmak amacıyla her bir modele dayalı elde edilen standart hata değerleri, ilişkili örneklemler için tek yönlü varyans analizi ile SPSS paket programı kullanılarak karşılaştırılmıştır (Pallant, 2007). Varyans analizi sonucunda modeller arasında manidar bir farklılık gözlenmiş ve bu doğrultuda alt testler için manidar farklılığın yönünü belirlemek amacıyla standart hata ortalamaları incelenmiştir.

Bulgular

Araştırmada öncelikle 1PLM, 2PLM, 3PLM ve STM'ye ilişkin model uyum istatistikleri verilerek yorumlanmış, ardından iki kategorili MTK modellerinin uyum düzeylerinin -2 loglikelihood değerlerine dayalı olarak varyans analizi ile ikili karşılaştırılmasına yönelik bulgular sunularak bu modellerin uyum düzeyleri değerlendirilmiştir. Sonrasında ise her bir model altında madde bazında uyumu değerlendirmek üzere hesaplanan S-X^2 istatistikleri incelenmiştir.

Model-veri uyumu modelin genel uyumu ve madde bazında uyum olarak ele alınmıştır ve bu doğrultuda öncelikle modellerin genel uyum düzeylerini değerlendirmek amacıyla modellere ilişkin uyum istatistikleri aşağıda Tablo 1'de sunulmuştur.

Tablo 1. Model Veri Uyumuna İlişkin Hesaplanan Uyum İstatistikleri

	RMSEA	TLI	CFI	SRMRS
1PLM	0.072317	0.949685	0.949949	0.123093
2PLM	0.020914	0.995792	0.996235	0.034145
3PLM	0.016637	0.997337	0.997898	0.022987
STM	0.022834	0.976346	0.985808	

Tablo 1 incelendiğinde modellere ilişkin hesaplanan RMSEA değerleri bir miktar farklılık sergilemekle birlikte diğer uyum indekslerinin modeller arası birbirine çok yakın olduğu ve tüm modellerin veriye iyi uyum sergilediği anlaşılmaktadır. Tüm modeller bir arada değerlendirildiğinde bu uyum istatistiklerine dayalı olarak model uyumu sıralaması 3PLM, 2PLM, STM ve 1PLM şeklindedir. RMSEA değerlerine bakıldığında 2PLM ve STM için hesaplanan değerler arası çok az fark olduğu görülmektedir. STM için SRMRS değerini sunulmasının nedeni ise, bu değerin sadece sıralı veriye dayalı olarak elde edilmesi ve dolayısıyla da sadece iki kategorili modelleri birbirleriyle karşılaştırmak amacıyla kullanılmasıdır (Chalmers, 2012). Bireysel olarak model uyum düzeylerinin değerlendirilmesi sonrasında ise modellere ilişkin -2 loglikelihood değerlerine dayalı olarak iki kategorili modellerin uyum düzeyleri varyans analizi yoluyla karşılaştırılmış ve analiz sonuçları Tablo 2'de sunulmuştur.

Tablo 2. İki Kategorili Modellerin Uyum Düzeylerinin Karşılaştırılmasına Yönelik Varyans Analizi Sonuçları

	1PLM	**2PLM**	**3PLM**
1PLM	(-16666.3)	0.00 (p<0.01)	0.00 (p<0.01)
2PLM	(-16123.6)		0.00 (p<0.01)
3PLM	(-15632.9)		

*p<.01

Tablo 2'de sunulan -2loglikelihood değerlerine ilişkin olasılık değerleri incelendiğinde, 1PLM'in veriye uyum düzeyi ile 2PLM'in uyum düzeyi arasında ve 2PLM'in veriye uyum düzeyi ile 3PLM'in uyum düzeyi arasında manidar farklılıkların bulunduğu görülmektedir. -2loglikelihood değerlerine bakıldığında ise 3PLM'in, 1PLM ve 2PLM'ye göre veriye daha iyi uyum sergilediği anlaşılmaktadır.

Bu karşılaştırmada STM'nin yer almamasının nedeni ise, STM ve iki kategorili MTK modellerine ilişkin veri setlerinin farklı olmasıdır. Genel model uyumunun değerlendirilmesi ve ikili

olarak karşılaştırılmasının ardından madde bazında uyumu değerlendirmek üzere hesaplanan madde uyum S-X^2 istatistikleri Tablo 3'te sunulmuştur.

Tablo 3. Madde Uyumuna İlişkin Hesaplanan S-X^2 İstatistikleri

Madde	1PLM	2PLM	3PLM	STM	Madde	1PLM	2PLM	3PLM	STM
m1	**0.0099**	0.2925	0.1575	0.6992	m11	**0.0006**	0.0375	0.4227	**0.0058**
m2	**0**	0.0395	0.7926	0.2819	m12	**0.0007**	0.0187	0.2197	**0**
m3	**0**	0.7971	0.9474	0.9356	m13	**0.0001**	**0.0007**	0.3034	0.0592
m4	0.1113	0.1024	0.1254	0.1148	m14	**0**	0.1232	0.1824	0.0195
m5	**0**	0.1786	0.5943	0.0706	m15	0.2323	0.5297	0.7054	0.107
m6	0.0972	0.4014	0.3103	0.0333	m16	**0**	0.0365	0.0311	0.5488
m7	**0.0038**	0.4081	0.2424	0.0292	m17	**0.0003**	0.109	0.1957	0.0935
m8	0.0904	0.0963	0.1011	0.1531	m18	**0.0002**	0.0473	0.0235	0.0102
m9	0.0516	0.9618	0.905	0.0149	m19	0.0112	0.282	0.1117	0.3778
m10	**0**	0.0144	0.0348	0.1528	m20	**0**	**0.0001**	0.0159	0.6384

Not: Anlamlı olan p değerleri koyu yazılmıştır.

Tablo 3'te sunulan değerler incelendiğinde 3PLM altında uyum sergilemeyen hiçbir maddenin bulunmadığı anlaşılmaktadır. 2PLM ve STM altında uyum sergilemeyen madde sayısı (2) eşittir. 1PLM'e dayalı olarak model testinde ise 14 maddenin veriye uyum sergilemediği görülmektedir. 2PLM ve STM altında uyum sergilemeyen madde sayısı eşit olmasına rağmen bu maddelerin farklı maddeler olduğu görülmektedir. STM altında 11. ve 12. maddeler uyum sergilemezken, 2PLM altında 13. ve 20. maddelerin uyum sergilemediği gözlenmiştir. Genel model uyumuna yönelik uyum istatistikleri ve madde bazında uyum istatistiklerine dayalı olarak bu veri seti için en iyi uyum sergileyen modelin 3PLM olduğu sonucuna varılmıştır.

Model veri uyumunun değerlendirilmesinin ardından araştırmada incelenen tüm MTK modelleri altında kestirilen madde parametreleri incelenmiştir. Ancak iki kategorili MTK modelleri ile STM'de kestirilen parametrelerin anlamları ve ele alınan kategori sayısı farklılaşmaktadır. STM altında kestirilen zeta (ζ) ve lambda (λ) parametreleri doğrudan yorumlanmak yerine, bu parametrelere doğrusal dönüşüm uygulanarak bu parametreler MTK'da iyi bilinen a ve b parametrelerine dönüştürülmektedir. (Ostini & Nering, 2006). Bu doğrultuda aşağıda öncelikle, Tablo 4'te iki kategorili MTK modelleri altında kestirilen madde parametreleri ve bu parametre kestirimlerine ilişkin standart hata değerleri sunularak bu modeller madde parametre kestirimleri açısından kendi aralarında karşılaştırılmıştır. Sonrasında ise STM altında madde parametre kestirimleri ve bu parametre kestirimlerine ilişkin standart hata değerleri verilerek yorumlanmıştır.

Tablo 4. İki kategorili MTK Modelleri Altında Kestirilen Madde Parametreleri ve Madde Parametre Kestirimlerine İlişkin Standart Hata (SH) Değerleri

	1PLM		**2PLM**				**3PLM**					
	b	**SH**	**b**	**SH**	**a**	**SH**	**b**	**SH**	**a**	**SH**	**c**	**SH**
m1	0.08	0.06	0.06	0.05	1.21	0.07	0.76	0.06	2	0.17	0.25	0.01
m2	0.24	0.06	0.05	0.03	2	0.1	0.18	0.04	2	0.11	0.04	0.01
m3	0.1	0.06	-0.03	0.03	2	0.1	0.09	0.04	2	0.11	0.04	0.01
m4	-0.83	0.06	-0.58	0.04	1.62	0.1	-0.05	0.05	2	0.16	0.29	0.02
m5	0.65	0.06	0.56	0.06	1.02	0.06	1.2	0.06	2	0.19	0.21	0.01
m6	0.35	0.06	0.16	0.04	1.75	0.09	0.53	0.05	2	0.13	0.13	0.01
m7	-0.1	0.06	-0.08	0.05	1.12	0.07	0.76	0.06	2	0.18	0.3	0.02
m8	0.45	0.06	0.2	0.04	1.88	0.09	0.47	0.04	2	0.12	0.08	0.01
m9	0.57	0.06	0.27	0.04	1.96	0.09	0.52	0.04	2	0.12	0.07	0.01
m10	1.59	0.07	2	0.12	0.52	0.03	2	0.12	0.5	0.03	0	0.01
m11	0.28	0.06	0.08	0.03	2	0.1	0.42	0.05	2	0.13	0.11	0.01
m12	0.24	0.06	0.08	0.04	1.81	0.09	0.49	0.05	2	0.14	0.15	0.01
m13	0.38	0.06	0.17	0.04	1.76	0.09	0.58	0.05	2	0.13	0.13	0.01
m14	0.57	0.06	0.49	0.06	1.04	0.06	1.14	0.06	2	0.18	0.21	0.01
m15	-0.4	0.06	-0.31	0.04	1.7	0.1	0.14	0.05	2	0.15	0.22	0.02
m16	-0.38	0.06	-0.31	0.03	2	0.11	-0.07	0.04	2	0.13	0.13	0.02
m17	0.12	0.06	-0.01	0.03	1.93	0.1	0.36	0.05	2	0.13	0.14	0.01
m18	0.87	0.07	0.45	0.04	2	0.09	0.67	0.04	2	0.11	0.04	0.01
m19	-0.17	0.06	-0.18	0.03	1.9	0.11	0.11	0.04	2	0.13	0.13	0.02
m20	1.24	0.07	2	0.13	0.46	0.03	2	0.13	0.46	0.03	0	0.02
SH_{ort}		0.062		0.049		0.084		0.056		0.129		0.013

Her bir modele dayalı olarak elde edilen madde güçlük parametresi (b) kestirimlerine bakıldığında genel olarak, 1PLM altında maddelerin orta güçlükte olduğu ancak 3PLM altında maddelerin güçlük düzeylerinin yükseldiği anlaşılmaktadır. Madde ayırt edicilik (a) kestirimleri incelendiğinde ise 2PLM altında daha düşük iken 3PLM altında yükseldiği anlaşılmaktadır. Bu tablonun ortaya çıkmasında c parametresinin etkili olduğu açıktır. Bu bulgu 3PLM'de maddelerin tamamının veriye uyum sergilemesi ile tutarlılık sergilemektedir. Şans parametresinin dikkate alınmasıyla 3PLM'de maddelerin ayırt edicilik düzeylerinin ve model-veri uyumunun artması beklenen bir durumdur. Tablo 4'te verilen parametre kestirimlerine ilişkin standart hata değerleri dikkate alındığında b parametresinin en az hatayla sırasıyla 2PLM, 3PLM ve 1PLM altında kestirildiği; ai parametresinin ise en az hatayla 2PLM altında kestirildiği anlaşılmaktadır.

İki kategorili MTK modellerine dayalı olarak madde parametre kestirimlerinin incelenmesinin ardından, STM altında kestirilen lambda ve zeta parametrelerine doğrusal dönüşüm uygulanarak a ve b parametre değerleri elde edilmiş ve Tablo 5'te sunulmuştur.

Tablo 5. STM Altında Kestirilen ve b Parametreleri

	a1	a2	a3	a4	b1	b2	b3	b4
M1	-0.72	-0.46	0.19	0.98	-0.4	-0.05	-0.28	0.73
M2	-0.73	-0.9	-0.37	2	-0.1	-0.14	-0.45	0.68
M3	-0.5	-0.9	-0.52	1.92	-0.61	-0.04	-0.17	0.82
M4	-0.5	-0.31	-0.54	1.35	-0.65	-0.14	-0.68	1.47
M5	-0.49	-0.07	-0.23	0.8	-0.37	-0.31	0.31	0.36
M6	-0.75	-0.48	-0.16	1.39	-0.53	-0.09	0.09	0.53
M7	-0.87	-0.43	0.2	1.09	-0.64	-0.88	0.52	1
M8	-0.66	-0.89	0	1.55	-0.04	-0.44	0.03	0.46
M9	-0.71	-0.45	-0.33	1.49	0.01	-0.19	-0.14	0.32
M10	-0.81	-0.49	0.74	0.56	-0.49	-0.48	0.78	0.2
M11	-0.8	-0.56	-0.18	1.54	-0.38	0.01	-0.22	0.59
M12	-0.75	-0.54	-0.16	1.45	-0.65	-0.11	0.12	0.64
M13	-0.73	-0.5	-0.18	1.41	-0.25	-0.4	0.14	0.51
M14	-0.28	-0.23	-0.29	0.8	-0.31	-0.22	0.15	0.38
M15	-0.46	-0.6	-0.3	1.36	-0.31	-0.49	-0.32	1.12
M16	-0.59	-0.68	-0.49	1.76	-0.55	-0.2	-0.47	1.22
M17	-0.39	-0.71	-0.35	1.46	-0.02	-0.24	-0.46	0.72
M18	-0.51	-0.62	-0.49	1.63	-0.22	-0.68	0.75	0.15
M19	-0.8	-0.52	-0.22	1.54	-0.37	-0.69	0.05	1.01
M20	-0.45	-0.06	0.19	0.33	-0.14	0.06	-0.03	0.11

(**Not:** a ve b parametreleri Multilog programında kestirildiğinden 4. kategori en yüksek kategori olarak ele alınmaktadır. Dolayısıyla bu parametrelere ilişkin 4. kategori doğru yanıtı göstermektedir.)

STM'de b parametresi yer parametresi olarak ele alınmaktadır. Tablo 5'te sunulan b parametresine ilişkin kategori değerleri doğru yanıt kategorisi (en yüksek kategori) bazında ele alınmıştır. Doğru yanıt kategorisine ilişkin b değerlerine bakıldığında, 4. maddede bu parametrenin değerinin en yüksek ve 20. maddede en düşük olduğu görülmektedir. Çeldiriciler açısından a ve b parametreleri kategori karakteristik eğrileriyle birlikte değerlendirilmiştir ancak yer sınırlılığı nedeniyle burada kategori karakteristik eğrileri sunulmamıştır. Bu inceleme sonucunda genel olarak, çeldiricilerin düşük yetenek düzeyindeki bireyler tarafından işaretlendiği ancak bazı maddeler için yetenek ölçeği boyunca (-3, 3) farklılaşmaların olduğu anlaşılmaktadır. Bu bağlamda en dikkat çekici bulgulardan biri, yetenek ölçeği üzerinde (-3, 1) aralığında birinci maddenin 3. ve 4. (doğru yanıt) kategorilerine ve ilişkin işaretlenme olasılıklarının ve kategori karakteristik eğrilerinin eğimlerinin birbirine yakın olmasıdır. Bu doğrultuda bu yetenek düzeyindeki bireyler için, birinci maddede doğru yanıt ile söz konusu çeldiricinin (3. kategori) benzer şekilde çalıştığı söylenebilir. Bununla birlikte onuncu madde için ise 3. yanıt kategorisi 4. yanıt (doğru yanıt) kategorilerinin işaretlenme olasılıklarının yetenek düzeyinin artmasına bağlı olarak artış gösterdiği yani yetenek ölçeğinde üst yetenek düzeyine doğru ilerlendikçe bu çeldiricinin doğru yanıt gibi çalıştığı anlaşılmaktadır. Son olarak ise tüm yetenek ölçeği boyunca (-3,3) on dördüncü madde için 1. ve 2. kategoriler ile on altıncı madde için 1. ve 3. kategorilerin eğrilerine ilişkin eğimlerdeki ve işaretlenme olasılıklarındaki değişimlerin paralellik gösterdiği gözlenmiştir. Düşük ve orta yetenek düzeyinde ((-3,0) yetenek aralığı için) olarak değerlendirilebilecek bireyler için

ise 20. maddenin 3. ve 4. kategorilerine ilişkin eğrilerin eğimlerinin benzer olduğu ancak yetenek ölçeğinde üst düzeye doğru gidildikçe bu kategorilerin işaretlenme olasılıkları ile kategorilerine ilişkin eğrilerin eğimlerinin arttığı anlaşılmaktadır.

Araştırmada incelenen bir diğer husus ise iki kategorili MTK modelleri ile STM altında kestirilen yetenek parametrelerinin doğruluk düzeyleridir. Bu doğrultuda her bir model altında yetenek parametresinin kestirimleri elde edilmiş ve bu kestirimlere ilişkin hata değerleri hesaplanarak karşılaştırılmıştır. Bu modellere dayalı olarak elde edilen yetenek parametre kestirimleri ile bu kestirimlere ilişkin standart hata değerlerine yönelik Betimsel İstatistikler hesaplanarak Tablo 6'da sunulmuştur.

Tablo 6.Yetenek Parametre Kestirimine İlişkin Standart Hatalara Yönelik Betimsel İstatistikler

Modeller	**N**	$\bar{X}$	**SS**	**Çarpıklık**	**Basıklık**
1PLM	1500	0.57	0.11	2.44	7.50
2PLM	1500	0.34	0.16	16.70	45.98
3PLM	1500	0.36	0.24	3.31	44.12
STM	1500	0.32	0.19	3.41	16.69

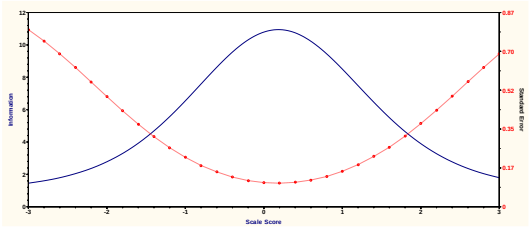
Tablo 6 incelendiğinde yetenek parametresinin en düşük hata ile STM altında kestirildiği anlaşılmaktadır. Yetenek parametresinin kestirimlerine ilişkin standart hata değerlerinin modeller arası manidar düzeyde farklılaşıp farklılaşmadığını belirlemek üzere uygulanan ilişkili örneklemler için tek faktörlü varyans analizi sonuçları, bu dört model altında yetenek kestirimlerine ilişkin standart hata değerleri arasında manidar düzeyde farklılık olduğunu ortaya koymuştur ($F(3, 4497) = 791{,}80$; $p<0{,}05$). Ardından yapılan ikili karşılaştırmalara dayalı olarak; olası bütün model çiftleri için bu farklılıkların manidar düzeyde olduğu, STM ile lojistik modeller arası yapılan karşılaştırmalarda ise STM'nin lehine bir farklılığın bulunduğu, 2PLM'nin 1PLM ve 3PLM'ye göre daha doğru, 3PLM'nin ise 1PLM'ye göre daha doğru yetenek kestirimleri sağladığı bulunmuştur.

Bu araştırmanın önemli bir diğer boyutu ise, düşük -orta ve yüksek yetenek düzeyinde her bir model altında test bilgi fonksiyonlarının karşılaştırılmalı olarak incelenmesidir. Tablo 7'de araştırmada incelenen tüm modellere dayalı olarak üç farklı yetenek düzeyinde bilgi fonksiyonunun sağladığı bilgi miktarı sunulmuştur:

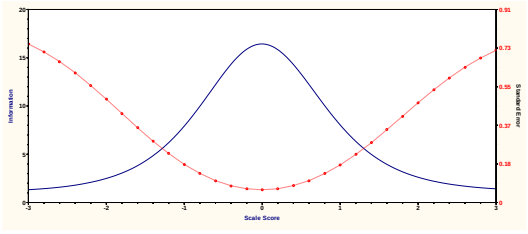
Tablo 1. Test Bilgi Fonksiyonları

Model/θ ölçeği yetenek aralıkları	**[-3, -1]**	**(-1, 1)**	**[1, 3]**	**[-3,3]**
1PLM	29.479	103.238	40.2	172.917
2PLM	28.617	137.116	29.74	195.473
3PLM	10.338	290.793	22.928	324.059
STM	37.06	148.425	32.592	218.077

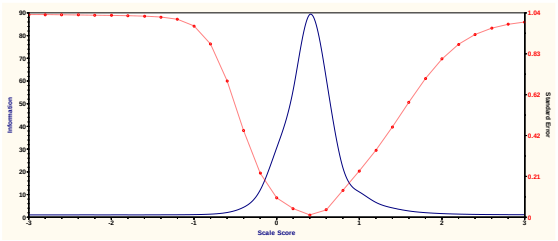
Tablo 7 incelendiğinde, SBS-2012'de uygulanan Matematik alt testinin, seviye belirleme amacıyla geliştirilmesini destekleyecek şekilde, hem iki kategorili modeller hem de STM altında yetenek kestirimi açısından en fazla bilgiyi orta yetenek düzeyindeki bireyler için sağladığı anlaşılmaktadır. Bununla birlikte düşük yetenek düzeyindeki bireyler için en fazla bilginin STM altında, orta yetenek düzeyindeki bireyler için en fazla bilginin 3PLM altında ve yüksek yetenek düzeyindeki bireyler için ise en fazla bilginin 1PLM altında sağlandığı gözlenmiştir. Tüm yetenek ölçeği (θ=-3, θ=+3) boyunca sağladığı bilgi miktarına bakıldığında, testin yetenek kestirimi konusunda en fazla bilgiyi 3PLM altında, en az bilgiyi ise 1PLM altında sağladığı anlaşılmaktadır. Testin bu üç yetenek düzeyinde test bilgi fonksiyonlarının sağladığı bilgi miktarının yanısıra, her bir model altında bu bilgi miktarının yetenek ölçeği boyunca değişimini gösteren test bilgi fonksiyonuna ilişkin grafikler de incelenmiştir. Bu doğrultuda aşağıda Şekil 1, Şekil 1, Şekil 3 ve Şekil 4'te her bir model altındaki test bilgi fonksiyonuna ilişkin grafikler sunulmuştur:



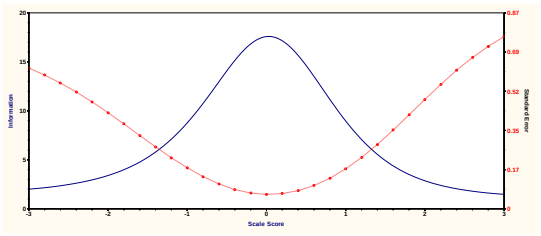
Şekil 1. 1PLM'ye Dayalı Test Bilgi Fonksiyonu Grafiği Grafiği



Şekil 2. 2PLM'ye Dayalı Test Bilgi Fonksiyonu



Şekil 3. 3PLM'ye Dayalı Test Bilgi Fonksiyonu Grafiği Grafiği



Şekil 4. STM'ye Dayalı Test Bilgi Fonksiyonu

Bu grafiklerde mavi ile çizilen eğriler, testin sağladığı bilgi miktarına kırmızı ile çizilen eğriler ise standart hataya ilişkindir. Bu doğrultuda her bir model altında hesaplanan fonksiyona ilişkin oluşturulan eğrinin tepe noktasının, -1 ile +1 aralığında yer aldığı anlaşılmaktadır. Bu aralık, testin en fazla bilgi sağladığı orta yetenek düzeyine denk gelmektedir. Grafikler incelendiğinde en fazla bilginin 3PLM altında sağlandığı ve bilginin maksimum olduğu noktada standart hata değerinin de en düşük değeri aldığı görülmektedir. Bununla birlikte 3PLM altında testin bilgi sağladığı yetenek ranjının daha dar olduğu anlaşılmaktadır. 1PLM altında elde edilen fonksiyon grafiği daha basık bir dağılımı gösterirken, 3PLM'e doğru gidildikçe daha sivri bir dağılım gözlenmekte ve aynı zamanda maksimum bilgi sağlanan yetenek aralığı da daralmaktadır.

Tartışma, Sonuç ve Öneriler

Bu çalışma, tek boyutlu iki kategorili MTK modelleri (1PLM, 2PLM 3PLM) ile çoklu puanlanan MTK modellerinden Sınıflamalı Tepki Modelinin model-veri uyumu, yetenek parametre kestirimlerinin doğruluğu ve test bilgi fonksiyonlarının sağladığı bilgi miktarı açısından karşılaştırılması amacıyla gerçekleştirilmiştir. Bu doğrultuda çoktan seçmeli test maddelerine verilen tepkileri modellemede ilgilenilen nitelikler açısından en iyi performans sergileyen model belirlenmeye çalışılmıştır.

Model-veri uyumu açısından bakıldığında, genel model uyumuna işaret eden uyum istatistiklerine dayalı olarak tüm modellerin veriye iyi uyum sergilediği anlaşılmaktadır. Model uyum düzeyleri karşılaştırıldığında ise çoktan seçmeli test maddelerine verilen tepkileri modellemede daha fazla parametreyi dikkate alan 3PLM'in diğer lojistik modellere göre veriye daha iyi uyum gösterdiği sonucuna varılmıştır. Bu bulgu Teker, Kelecioğlu ve Eroğlu (2013) ile Önder (2007)'in çalışmalarından elde edilen bulgular ile tutarlılık sergilemektedir. Araştırmacılar, iki kategorili MTK modellerinin genel uyum düzeylerinin, -2 loglikelihood değerlerine göre karşılaştırılması sonucunda 3PLM' in en iyi uyum sergileyen model olduğunu rapor etmişlerdir. Yapılan bir başka çalışmada (Çelik, 2001) ise iki kategorili MTK modellerinin matematik alt testinden elde edilen veriye uyum düzeylerinin, ki kare istatistiklerine göre karşılaştırılması sonucunda da benzer bulgular elde edilmiştir. Model-veri uyumuna ilişkin tüm bulgular (genel model uyumu ve madde bazında uyuma ilişkin) değerlendirildiğinde en iyi uyum gösteren modellerin sırasıyla 3PLM, STM,2PLM ve 1PLM olduğu sonucuna ulaşılmıştır.

Genel uyumun değerlendirilmesi sonrasında model uyumu, madde bazında değerlendirilmiştir. İki kategorili MTK modelleri için, kestirilen parametre sayısı arttıkça madde uyumunun artış sergilediği gözlenmiştir. 3PLM altında tüm maddeler iyi uyum sergilerken, STM VE 2PLM altında 18 maddenin veriye iyi uyum sergilediği, 1PLM altında ise az sayıda (6) maddenin veriye uyum sergilediği bulunmuştur. Bu bulgu ile tutarlı olarak Önder (2007) gerçekleştirdiği çalışmaya dayalı olarak fazla maddenin 3PLM altında ve en az maddenin ise 1PLM altında uyum sergilediğini ifade belirtmiştir. Bunun yanısıra hali hazırdaki çalışmada 2PLM ve STM altında, 1PLM altında uyum gösteren 6 maddeyi de kapsayacak şekilde 18 maddenin uyum göstermesi, söz konusu maddeler için ayırt edicilik parametrelerinin de dikkate alınmasının önemli olduğunu göstermektedir. Hem madde bazında uyum hem de genel model uyumu açısından bakıldığında 2PLM ve STM' nin benzer sonuçlar ürettiği görülmektedir. Bu iki MTK modelinde dikkate alınan parametre türünün benzer olması (a ve b parametreleri), ve parametre türünün de model test sonuçlarında etkili bir faktör olması nedeniyle bu tablonun ortaya çıktığı düşünülmektedir.

Hali hazırdaki çalışmanın sonuçları model seçiminin, yetenek parametresinin kestiriminin doğruluğunu etkilediğini ortaya koymaktadır. Çoktan seçmeli test maddeleri için STM 'ye dayalı olarak, iki kategorili MTK modellerine göre daha doğru yetenek kestirimleri elde edilmiştir. Bu bulgular, Bock (1972)'ın ve Ching-Fung (2002)'un çalışmasında elde ettiği bulgular ile benzerlik göstermektedir. Bock (1972) yaptığı çalışma sonucunda orta yetenek düzeyinin altındaki bireyler için, çoklu puanlanan MTK modelleri altında, iki kategorili MTK modellerine göre önemli ölçüde daha az hatayla yetenek parametresinin kestirildiğini belirtmiştir. Ching-Fung (2002) da benzer şekilde, çoklu puanlanan MTK modellerinin iki kategorili MTK modellerine göre daha doğru yetenek kestirimleri ile sonuçlandığını gözlemlemiştir. Ayrıca DeAyala ve Koch (1987) çalışmasında CAT uygulamalarına dayalı olarak STM altında ve 3PLM altında elde ettiği yetenek kestirimlerinin, STM'de daha düşük standart hata ile sonuçlandığını gözlemlemiştir. Bu çalışmadan elde edilen bulgular da araştırma sonuçlarını desteklemektedir. Bu çalışma ile bu konuda yapılan diğer çalışmaların (Bock, 1972; Ching-Fung, 2002; DeAyala & Koch, 1987) sonuçları; çoktan seçmeli test maddelerine verilen tepkilere dayalı olarak yetenek kestiriminde yanıt kategorilerine verilen tepkilerin dikkate alınması gerektiğini göstermektedir.

Bu çalışmanın diğer bir boyutunu ise düşük-orta ve yüksek yetenek düzeyindeki bireyler için test bilgi fonksiyonlarının sağladığı bilgi miktarının her bir model altında incelenmesi oluşturmaktadır. Bu bağlamda elde edilen bulgular, Matematik alt testinde en fazla bilgiyi sağlayan modellerin sırasıyla 3PLM, STM, 2PLM ve 1PLM olduğunu ortaya koymaktadır. Testin bilgi miktarı ise standart hata ile ilişkilidir ve standart hata değerlerinin düşük olması, testin sağladığı bilgi miktarının artmasını sağlamaktadır (Hambleton & Swaminathan, 1985, s.25-50). Test bilgi fonksiyonlarının sağladığı bilgi miktarı ve yetenek parametre kestirimlerine ilişkin standart hata değerlerinin birlikte değerlendirmesine dayalı olarak, STM'nin 1PLM ve 2PLM'ye göre daha yüksek test bilgi fonksiyonları sağladığı ve yetenek parametresinin daha az hatayla kestirildiği sonucuna ulaşılmıştır. Bununla birlikte STM altında yetenek parametresi 3PLM'ye göre daha az hatayla kestirilmesine rağmen, 3PLM'de test bilgi fonksiyonun sağladığı bilgi miktarının STM'den daha yüksek olduğu dikkat çekmektedir. Bu durumun, madde ayırt edicilik parametreleri ile ilişkili olabileceği düşünülmektedir. 3PLM altında maddelerin büyük bir çoğunluğu için madde ayırt edicilik parametre kestirimlerinin maksimum değerlerine ulaştığı ve dolasıyla da maddelerin sağladığı bilgi miktarının da çok yüksek değerlere ulaştığı görülmüştür. Benzer şekilde Hambleton ve Jones (1994) da yaptıkları çalışmaya dayalı olarak ayırt ediciliği yüksek olan maddeler için madde bilgi miktarının olduğundan yüksek kestirildiğini rapor etmişlerdir.

Bu durumun ayrıca ayırt edicilik parametresi ile yetenek ölçeği üzerinde ele alınan bölge arasındaki ilişkisi bağlamında da açıklanabileceği düşünülmektedir. Madde ayırt edicilik parametresinin kestirimi arttığında ilgili maddenin yetenek ölçeği üzerinde farklı noktalarda yer alan bireyleri ayırma gücü artarken, maddenin yetenek ölçeği üzerinde ayırım yaptığı yetenek ranjı daralmaktadır (Harris, 1989). Embretson ve Reise (2000)'e göre yüksek ayırt ediciliğe sahip madde, tepe yapmış bir bilgi eğrisi

sunar ve bu durumda madde dar bir yetenek aralığında çok bilgi sağlar. Bunun aksine ayırt ediciliği düşük düzeyde olan maddeler için ise bilgi fonksiyonuna ilişkin eğri daha düz olur ve daha geniş bir yetenek ranjını kapsar.

Test bilgi fonksiyonlarına ait grafikler incelendiğinde 1PLM' den 3PLM' ye doğru fonksiyon eğrisinin daraldığı, bireyin yeteneğine ilişkin en fazla bilginin 3PLM altında elde edilmesine rağmen, bu model altında bilgi sağlanan yetenek aralığının diğer modellere göre oldukça dar olduğu görülmektedir. Bu konuda yapılan çalışmalar (DeAyala, 1989; Borgatto, Azevedo, Pinherio & Andrade, 2015), kuramsal bilgilerle doğrudan örtüşmemesine rağmen, test bilgi fonksiyonunun yetenek ölçeği üzerinde düşük olduğu durumda daha doğru yetenek kestirimlerin yapıldığını göstermektedir. Borgatto, Azevedo, Pinherio ve Andrade (2015) gerçekleştirdikleri çalışmaya dayalı olarak, yetenek ölçeği üzerinde en az bilgi elde edilen yetenek aralığında daha az hatayla yetenek kestirimlerinin yapıldığını bildirmişlerdir.

DeAyala (1989) çalışmasında STM'ye dayalı olarak test bilgi fonksiyonunu hesaplamış ve bu fonksiyona ilişkin grafiğin pozitif kayışlı olduğunu belirtmiştir. Bu bulgu STM'nin, üst yetenek düzeyindeki bireyler için daha fazla bilgi sağladığı anlamına gelmektedir. STM ve 3PLM altında test bilgi fonksiyonlarının sağladığı bilgi miktarı karşılaştırıldığında, orta yetenek düzeyi için bu iki model altında sağlanan bilgi miktarının benzer olduğu sonucuna ulaşılmıştır. Bu bulgular Borgatto ve diğerleri'nin (2015) ve DeAyala (1989)'nın çalışmalarından elde edilen bulgular ile de desteklenmektedir.

Bu araştırma ile bu konuda yapılmış olan tüm bu çalışmaların sonuçları ve ilgili alandaki kuramsal açıklamalar, çoktan seçmeli test maddelerine dayalı olarak bireylerin yeteneklerinin kestiriminde, tüm yanıt kategoriline verilen tepkileri dikkate alan çoklu puanlanan MTK modellerinden STM'nin kullanılmasının bireylerin kısmı bilgisinin de değerlendirilmesini sağlayarak daha doğru yetenek kestirimlerinin elde edildiğini göstermektedir. Çoktan seçmeli test maddeleri, bireyler hakkında birçok alanda önemli kararlar vermeye yönelik yapılan ulusal ve uluslararası geniş ölçekli sınavlarda, yaygın bir şekilde kullanılmaktadır. Ancak KTK çerçevesinde puanlama yapılması ya da yetenek kestirimlerinin iki kategorili MTK modellerine dayalı olarak elde edilmesi, çoktan seçmeli test maddelerindeki çeldiricilerin puanlamaya dahil edilmemesine ve dolayısıyla da bilgi kaybına neden olabilmektedir. Bu nedenle, STM birçok alanda sıklıkla kullanılan çoktan seçmeli test maddelerinin değerlendirilmesinde daha ayrıntılı bilgiler sağlayacağından uygulamaya katkısı olacağı düşünülmektedir. Ayrıca, STM altında maddelerin tepki kategorileri yetenek ölçeği boyunca incelenmekte ve çeldiricilerin işlevselliği ortaya konulabilmektedir. Bu doğrultuda çoktan seçmeli test maddelerinde önemli bir husus olan çeldiricilerin etkililiği incelenerek maddelerin her yönüyle test edilip geliştirilmesine katkıda bulunacaktır. Böylelikle, STM'nin uygulamada madde bankası oluşturma çalışmalarına katkı sağlaması beklenilmektedir.

Bu araştırmada STM ile iki kategorili MTK modelleri; model-veri uyumu, madde parametrelerinin kestirimi ve düşük-orta ve yüksek yetenek düzeylerindeki bireyler için testin sağladığı bilgi miktarı açısından karşılaştırılmıştır. Bu karşılaştırma için farklı yetenek grupları, araştırmada kullanılan SBS-2012 Matematik alt testinden elde edilen ölçümlere dayalı olarak belirlenmiştir. Ancak alanyazında (de la Torre & Hong, 2010), yanıtlayıcıların yetenek düzeylerinin test uygulaması öncesinde belirli ölçümlere dayalı olarak belirlenerek yetenek grupları oluşturulmasının, yetenek kestirimlerinin daha sağlam olmasına yol açacağı belirtilmektedir. Bu doğrultuda sonraki araştırmalarda, bireylerin yeteneklerinin önceki ölçümlere dayalı olarak belirlenip farklı yetenek gruplarının oluşturulması ve ardından farklı yetenek düzeyindeki bireyler için STM'nin etkililiğinin diğer MTK modelleri ile karşılaştırılması önerilebilir. Araştırmada incelenen her bir MTK modeli altında test bilgi fonksiyonunun sağladığı bilgi miktarı farklı yetenek düzeyleri için karşılaştırmalı olarak incelendiğinde, düşük yetenek düzeyindeki bireyler için en fazla bilgi sağlayan modelin STM olduğu, diğer yetenek düzeylerinde ise iki kategorili modellerden 3PLM'nin daha fazla bilgi sağladığı

gözlenmiştir. Bu nedenle STM'nin düşük yetenek düzeyindeki bireyler için kullanılması önerilmektedir. Ayrıca STM'nin farklı yetenek düzeyleri için performansı, iki kategorili MTK modellerinden 4-Parametreli Lojistik Model (4PLM) ve çoklu puanlanan MTK modelleri ile karşılaştırmalı olarak incelenebilir. STM, çoktan seçmeli test maddesindeki tüm yanıt kategorilerini dikkate aldığı için değişen çeldirici fonksiyonunu (differential distractor function) ortaya çıkarmak amacıyla kullanılabilmektedir (Penfield, 2008). Bu doğrultuda maddenin kategori bazında analizleri yapılacağından daha işlevsel maddeler geliştirilebilecektir. Madde havuzu oluşturmak amacıyla başarı testi maddelerinin STM ile analiz edilmesi önerilmektedir.

KAYNAKÇA

Baker, F. B. (1992). *Item response theory: parameter estimation technique.* New York: Marcel Dekker.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord and M. R. Novick, *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika, 37(1),* 29–51.

Bock, R. D. (1997). A brief history of item response theory. *Educational Measurement: Issues and Practice, 16(4),* 21–33.

Borgatto, A. F., Azevedo, C. L. N., Pinheiro, A., & Andrade, D. F. (2015). Comparison of ability estimation methods using irt for test with different degrees of difficulty. *Communications in Statistics – Simulation and Computation, 44(2),* 474-488.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.

Chalmers, R. P. (2012). mirt: A multidimensional ıtem response theory package for the R Environment. *Journal of Statistical Software, 48(6),* 1-29. URL http://www.jstatsoft.org/v48/i06/.

Ching-Fung, B. S. (2002). *Ability estimation under different item parametrization and scoring models.* Doctorate Dissertation, Nova Southeastern University, North Texas University, Texas,USA.

Çalışkan, M. (2000). *The Fit of one, two and three-parameter models of item response theory (IRT) to the ministry of National Education-Educational Research and Development Directorates's (MNE-ERDO) science achievement test data*. Yüksek Lisans Tezi, Orta Doğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.

Çelik, D. (2001). *The fit of one, two and three-parameter models of item response theory (IRT) to the ministry of National Education secondary school institutions student selection and placement test data.* Yüksek Lisans Tezi, Orta Doğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.

De la Torre, J., & Hong, Y. (2010). Parameter estimation with small sample size a higher-order irt model approach. *Applied Psychological Measurement, 34(4),* 267–285.

DeMars, C. E. (2008). *Scoring multiple choice ıtems: A comparison of IRT and classical polytomous and dichotomous methods.* Center for Assessment and Research Studies, James Madison University, Harrisonburg.

De Ayala, R.J. & Koch, W.R. (1987). Computerized adaptive testing: A comparison of the nominal response model and the three parameter model. *Paper Presented at Annual Meeting* of the *National Council on Measurement in Education Nisan 1987,* Washington, D.C, USA.

De Ayala, R.J. & Koch, W.R. (1988). Computerized adaptive testing: A comparison of the nominal response model and the three parameter model. *Paper Presented at Annual Meeting of American Research Association, 1988,* New Orleans, USA

De Ayala, R.I. (1989). Computerized adaptive testing: A comparison of the nominal response model and the three-parameter model. *Educational and Psychological Measurement, 49*(4), 789-805.

De Ayala, R. J. (1992). The nominal response model in computerized adaptive testing. *Applied Psychological Measurement, 16,* 327-343.

De Ayala, R. J. (1993). An introduction to polytomous item response theory models. *Measurement and Evaluation in Counseling and Development, 25*, 172- 189.

De Ayala, R. J. (2009). *The theory and practice of ıtem response theory*. U. S. A.

Demirtaşlı, N., Yalçın, S. & Ayan, C. (2014*).* Ölçme ve değerlendirme dersine yönelik tutum ölçeğinin çok kategorili madde tepki kuramı modellerine dayalı olarak ölçeklenmesi. *IV. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresinde Sunulmuş Bildiri [IV. International Congress on Measurement and Evaluation in Education and Psychology*] Hacettepe Üniversitesi, Ankara.

Dimitrov, D. (2003). Marginal true-score measures and reliability for binary items as a function of their IRT parameters. *Applied Psychological Measurement, 27*, 440-458.

Drasgow, F., Levine, M. V., Tsien, S., Williams, B., & Mead, A. D. (1995). Fitting polytomous item response models to multiple-choice tests. *Applied Psychological Measurement, 19,* 143-165.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. *Multivariate applications book series (Vol. 4).* New Jersey: Lawrence Erlbaum Associates.

Foley, B. P. (2010). *Improving IRT parameter estimates with small sample sizes: Evaluating the efficacy of a new data augmentation technique*. Doktora Tezi, The College of Education and Human Sciences, Nebraska.

Fraenkal, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education (8th ed.)*. New York: Mc Graw Hill.

Hambleton, R. K., & Swaminathan, H. (1985*). Item response theory: Principles and applications.* Boston: Kluwer.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of item response theory. *Contemporary Sociology* ,(21).

Hambleton, R. K., & Jones, R. W (1994). Item parameter estimation errors and their influence on test information functions. *Applied Measurement in Education, 7(3),* 171-186.

Harris, D. (1989). Comparison of 1-, 2-, and 3-Parameter IRT models. *Instructional Topics in Educational Measurement, 8(1),* 35-41.

Harvey, R. J., & Hammer, A. L. (1999). Item response theory. *Counseling Psychologist, 27*(3), 353–374.

Kılıç, İ. (1999). *The fit of one- two- and three- parameter models of item response theory to the student selection test of the student selection and placement center.* Doktora Tezi, Orta Doğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.

Lord, F.M. ve Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

MEB. (2009). *İlköğretim matematik dersi 6-8. sınıflar öğretim programı ve kılavuzu*. 15 Temmuz 2014 tarihinde http://ttkb.meb.gov.tr/www/ogretim-programlari/icerik/72 sayfasından erişilmiştir.

Mislevy, R. J. (1995). Foundations of a new test theory. In N. Fredericksen., R. J. Mislevy, & I. I. Bejar (Eds.), *Test theory for a new generation of tests (pp. 19-39)*. Hillsdale, NJ: Lawrence Erlbaum.

Ostini, R. & Nering, M. L. (2006). *Polytomous item response model*. U.S.A.

Osterlind, S. J. (1998). *Constructing test items: Multiple-choice, constructed response, performance, and other formats (2nd ed.).* Boston: Kluwer Academic.

Önder, İ. (2007). An investigation of goodness of model data fit. *Hacettepe University Journal of Education, 32,* 210-220.

Özkurt, S. (2002). *The Fit of one-, two- and three-parameter models of item response theory (IRT) to An English proficiency achievement test data.* Yüksek Lisans Tezi, Orta Doğu Teknik Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.

Pallant, J. (2007). *SPSS Survival Manual: A Step by Step Guide to Data Analysis using SPSS for Windows*. New York: Open University Press.

Penfield, R. D. (2008). An odds ratio approach for assessing differential distractor functioning effects under the nominal response model. *Journal of Educational Measurement, 45(3),* 247-269.

Penfield, R. D., de de la Torre (2008). A new response model for multiple-choice items. Paper presented at the 2008 annual meeting of the National Council on Measurement in Education, New York.

Stone, A. (2006). *A Psychometric analysis of the Statistics Concept Inventory*. Doctorate Dissertation, University of Okhloma, U. S. A.

Şahin, A. (2012). *Madde tepki kuramında test uzunluğu ve örneklem büyüklüğünün model veri uyumu, madde parametreleri ve standart hata değerlerine etkisinin incelenmesi*. Doktora Tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

Tay, L., Huang, Q., & Vermunt, J. K. (2016). Item response theory with covariates (IRT-C): Assessing item resovery and item functioning fort he three-parameter logistic model. *Educational and Psychological Measurement, 76(1),* 22–42.

Teker, G. T., Kelecioğlu H., & Eroğlu, M. G. (2013). An investigation of goodness of model data fit. *Social and Behavioral Sciences, 106,* 394-400.

Thissen, D. & Steinberg, L. (1986). Taxonomy of Item Response Models. *Psychometrika, 51(4),* 567-577.

Thissen, D., Steinberg, L. & Fitzpatrick, A.R. (1989.). Multiple- choice models: The distractors are also part of the item. *Journal of Educational Measurement, 26(2),* 161 – 176.

Tomkowicz, J. T. (2000). *Comparison of Ability Estimates From Dichotomously and Nominally Scored Testwise Susceptible and Non-Susceptible Items.* Doctoral Thesis. University of Alberta, Edmonton.

Yen, 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.