

# Tek ve Çok Boyutlu Madde Tepki Kuramına Dayalı Bir Veri Analizi Yazılımı: IRTPRO 2.1

## A Data Analysis Software Based on Uni- and Multidimensional Item Response Theory: IRTPRO 2.1

Esin YILMAZ KOĞAR \*

Derya ÇAKICI ESER \*\*

### Öz

Bu çalışmada, farklı madde tepki kuramı (MTK) modellerine göre analiz yapmaya olanak sağlayan “IRTPRO 2.1 for Windows” adlı bilgisayar programını tanıtmak amaçlanmıştır. Bu doğrultuda çalışmada IRTPRO 2.1 yazılımı için gerekli donanım özellikleri, programa erişim, programın çalıştırılması, analizler, çıktı dosyaları, uyum iyiliği indeksleri ve betimsel istatistikler üzerinde durulmuştur. Ayrıca farklı yapılarla ilişkin 20 madde ve 2000 kişilik simülatif verilerle örnek uygulama yapılmıştır. Elde edilen sonuçlardan yola çıkarak farklı IRT modellerini analiz etmek için gerekli kestirim süresi hakkında araştırmacılara bilgi vermek amaçlanmıştır. Yapılan derlemeler ve araştırmacıların deneyimleri sonucunda elde ettikleri bilgiler yardımıyla IRTPRO 2.1 hakkında bir ön bilgi verilmeye çalışılmış, programın avantajlı ve dezavantajlı yönleri tartışılmıştır. IRTPRO 2.1; MTK’ya dayalı veri analizi yapan programların yerine getirebildiği fonksiyonların tek bir programla yapılabilmesini sağlaması bakımından kullanışlı bir yazılımdır. Çalışmalarında bu yazılımdan yararlanmak isteyen araştırmacılara, üstünlükler ve sınırlılıklar açısından yazılımın deneme sürümünü incelemeleri ve bundan sonra lisansı satın alıp almamaya karar vermeleri önerilmektedir.

*Anahtar Kelimeler:* Anahtar Kelimeler: Madde tepki kuramı, veri analizi, IRTPRO 2.1, bilgisayar yazılımı

### Abstract

The purpose of this study is to introduce “IRTPRO 2.1 for Windows” computer software which enables to perform analysis based on different item response theory (IRT) models. In this context, the study is focused on the necessary hardware features for IRTPRO 2.1 software, access to the program, running the program, analysis, output files, the goodness of fit index and descriptive statistics. In addition, sample applications are made with simulative data belonging different structures including 20 items and 2000 examinees. Based on the obtained results it is intended to provide information about the amount of time needed to analyze different IRT models. As a result of compilations and experiences of researchers, a prior knowledge have tried to give about IRTPRO 2.1, the advantages and disadvantages of the program were discussed. IRTPRO 2.1 is a practical software and allows to execute functions, which can be performed by several softwares for data analysis based on IRT, with one program. Researchers who wish to make use this program in their research are recommended to investigate the trial version of the software in terms of strengths and limitations and then decide whether to buy a license.

*Keywords:* Item response theory, data analyses, IRTPRO 2.1, computer software

### GİRİŞ

Klasik Test Kuramı’nın (KTK) yeterli olmadığı noktalarda test uygulayıcılarına ve araştırmacılara çözümler sunmak amacıyla geliştirilen Madde Tepki Kuramı (MTK) test eşitleme, değişen madde

\* Araş. Gör., Hacettepe Üniversitesi, Eğitim Fakültesi, Ankara-Türkiye, esinyilmazz@gmail.com

\*\* Dr., Kırıkkale Üniversitesi, Eğitim Fakültesi, Kırıkkale-Türkiye, deryacakicieser@gmail.com

fonksiyonu ve bilgisayara dayalı bireyselleştirilmiş test uygulamalarında da KTK'ya karşı güçlü bir kuramdır. Bu kurama dayalı yapılacak uygulamalar için farklı algoritmaları kullanan birçok bilgisayar programları geliştirilmiştir. Uzun yıllardır MTK'ya dayalı madde ve yetenek parametresi kestirimleri ve farklı MTK uygulamaları için BILOG-MG (Zimowski, Muraki, Mislevy ve Bock, 2003), MULTILOG (Thissen, 2003), PARSCALE (Muraki ve Bock, 2003), WINSTEPS (Linacre, 2011) ve Xcalibre gibi yazılımlar kullanılmaktadır. Ayrıca açık kaynaklı ücretsiz bir yazılım olan R yazılımının ltm, eRm, plRasch, plink ve mirt (Rusch, Mair ve Hatzinger, 2013; Rizopoulos, 2015) paketlerinde tanımlanmış kodlar aracılığıyla çeşitli MTK uygulamaları gerçekleştirilebilmektedir. Bu yazılımlardan BILOG-MG, R'ın eRm ve plRasch paketleri iki kategorili verileri, MULTILOG ve R'ın plink paketi çok kategorili verileri analiz etmek için kullanılırken; PARSCALE, WINSTEPS, Xcalibre ve R'ın ltm paketi ile hem çok kategorili hem iki kategorili veriler analiz edilebilmektedir (Rizopoulos, 2015).

MTK, tek boyutluluk varsayımına dayanmaktadır. Bu nedenle bir veri seti üzerinde MTK'ya dayalı bir uygulama yapmak için ölçme aracının baskın bir faktörü ölçtüğünün gösterilmesi gereklidir (Hambleton ve Swaminoton, 1985; Embretson ve Reise, 2000). Ancak yapılan ölçme işlemlerinde tek boyutluluk varsayımı her zaman sağlanamayabilir ya da yapılan ölçme işlemi ile birden fazla özelliğin ölçülmesi amaçlanabilir. Bu sebeplerle, ortaya atıldığı dönemden 1970'lerin sonuna kadar tek boyutlu testler için kullanılan MTK, 1970'lerin sonu ve 1980'lerin başından itibaren çok boyutlu testlere genişletilmiş ve çok boyutlu madde tepki kuramı (ÇBMTK) başlığı altında birden fazla yeteneği ölçen testlerde de uygulanmaya başlanmıştır (Ansley ve Forsyth, 1985; Reckase, 2009). Çok boyutlu testlerde madde ve birey parametreleri tek boyutlu MTK'dan farklı biçimde; ÇBMTK'ya özgü eşitliklerinden yararlanarak kestirilmektedir. Bu sebeple ortaya konan yeni kuram ile yeni yazılımların geliştirilmesi ve kullanılması ihtiyacı doğmuştur.

Verileri ÇBMTK'ya dayalı olarak analiz etmek için kullanılan yazılımlardan TESTFACT (Bock vd., 2003), NOHARM (Fraser ve McDonald, 1988) BMIRT II (Yao, 2003), IRTPRO 2.1 (Cai Thissen ve du Toit, 2011) ve flexMIRT 2 (Cai, 2013) literatürde sıkça kullanılmaktadır. Ancak programlar çeşitli açılardan birbirinden farklılaşmaktadır. Buna göre NOHARM ve BMIRT ücretsiz, TESTFACT, IRTPRO 2.1 ve flexMIRT 2 ticari yazılımlardır. NOHARM ile madde parametrelerinden ayırt edicilik ve güçlük kestirilmekte, ancak yetenek parametresi kestirimi yapılamamaktadır. Buna karşılık TESTFACT, BMIRT, IRTPRO 2.1 ve flexMIRT 2'de hem madde hem de birey parametreleri analiz edilebilmektedir. Ayrıca programların parametre kestirimi için kullandığı yöntemler de farklılık göstermektedir. Çok boyutlu MTK'ya dayalı kestirim yapan programların IRTPRO 2.1 ile karşılaştırması Tablo 1'de özet olarak verilmiştir.

Tablo1. Çok Boyutlu MTK Kuramına Dayalı Veri Analizi Programlarının Karşılaştırılması

Özellikler	TESTFACT	NOHARM	BMIRT	FlexMIRT 2	IRTPRO 2.1
Şans parametresi kestirimi	✓	X	✓	✓	✓
Birey parametresi kestirimi	✓	X	✓	✓	✓
İki kategorili veriler	✓	✓	✓	✓	✓
Çok kategorili veriler	✓	X	✓	✓	✓
Profesyonel el kitabı	✓	X	X	✓	✓
Komut ile çalışma	✓	✓	✓	✓	✓
Tıklayarak (menüden seçim ile) çalışma	X	X	X	X	✓

Tablo 1'e göre NOHARM şans parametresi ile birey parametresini kestirememekte ve çok kategorili verileri analiz edememektedir. BMIRT'in ve NOHARM'ın profesyonel bir el kitabı bulunmamaktadır. Tabloda yer alan tüm yazılımlardan IRTPRO 2.1 dışında kalanlar yalnızca yazılan komut ile çalışmaktadır. Ancak IRTPRO 2.1 hem komut ile hem de menüden seçim yapılarak çalıştırılabilmektedir.

IRTPRO 2.1 (Item Response Theory for Patient-Reported Outcomes) programı Li Cai, David Thissen ve Stephen du Toit tarafından 2011 yılında geliştirilmiştir. IRTPRO madde kalibrasyonu ve test puanlama için MTK'yı kullanan istatistiksel bir yazılımdır. IRTPRO 2.1'de aşağıda maddeler halinde verilen MTK modelleri kullanılmaktadır (Cai, Thissen ve du Toit, 2011):

- İki parametrelili lojistik model (2PL) [eşit ayırt edicilik parametresi kullanılması ile bir parametrelili lojistik model haline gelir (1PL)]
- Üç parametrelili lojistik model (3PL)
- Aşamalı Tepki Modeli
- Genelleştirilmiş Kısmi Puan Modeli
- Adlandırmalı Tepki Modeli

Bu çalışmada, yukarıda belirtilen farklı madde tepki kuramı modellerine göre analiz yapmaya olanak sağlayan "IRTPRO 2.1 for Windows" adlı bilgisayar programını tanıtmak amaçlanmıştır. Bu amaçla çalışmada programın kurulumu, analizler, çıktılar ve kestirim süreleri hakkında genel bilgilere yer verilmiştir. IRTPRO 2.1 ile belirtilen modellere ilişkin hem tek boyutlu hem çok boyutlu testler analiz edilebilmektedir. Ayrıca bu yazılım ile araştırmacının amacına ve ölçme aracının yapısına bağlı olarak bu modellerin farklı kombinasyonlarının yer aldığı bir testin analizini yapmak mümkündür.

IRTPRO 2.1 yazılımı yurt dışında oldukça sık karşımıza çıkan bir yazılım olmasına karşılık, yurt içinde yapılan MTK ve ÇBMTK çalışmasında oldukça az kullanılmıştır. Bu durumun nedeni olarak IRTPRO 2.1 yazılımının yurt içinde yeteri kadar tanınmaması görülmektedir. IRTPRO 2.1'in tanıtılmasını amaçlayan herhangi bir yurt içi çalışmanın olmaması ve bu çalışmada MTK tabanlı çalışma yapmak isteyen araştırmacılara yardımcı olacak bilgilerin sunulması sebebiyle çalışmanın önem taşıdığı düşünülmektedir. Kullanım kolaylığına sahip bu programın tanıtımı ile özellikle MTK tabanlı çalışma yapmak isteyen araştırmacılara yardımcı olacak bilgiler sunulmaktadır. Bu bilgiler ışığında da araştırmacılar gelecekte yapacakları çalışmalarını planlayabileceklerdir. Araştırmanın bu bakımlardan alan yazına katkı getireceği düşünülmektedir.

## **PROGRAMIN KURULMASI, ANALİZLER, ÇIKTILAR ve KESTİRİM SÜRELERİ**

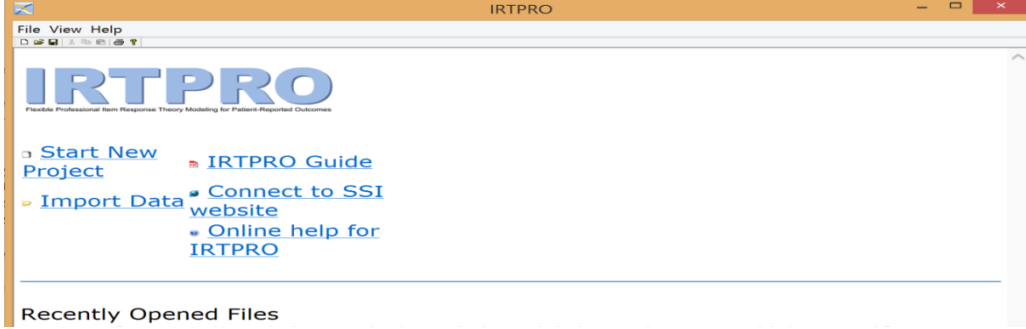
### ***Donanım Gereksinimleri ve Programa Ulaşım***

Program, Windows7, Vista ve XP işletim sistemleri ile kullanılabilir. Programa <http://www.ssicentral.com> adresinden ulaşılabilir. Yazılımın en fazla 25 madde, 1000 kişi ve 3 boyut ile çalışabilen öğrenci versiyonuna ve 15 günlük deneme sürümüne aynı adresten ücretsiz olarak erişilmektedir. Akademik olan ve akademik olmayan kullanıcılara yönelik altı aylık ve on iki aylık tam sürümlerin 2015 yılındaki ücreti 75\$ ile 950\$ arasında değişmektedir. Programın yazılım anahtarı, ücretin internet üzerinden ödenmesinden sonra elektronik posta yoluyla gönderilmektedir. Programın ücretli versiyonlarını kullanan kişilere teknik destek hizmeti de verilmektedir. Tam sürümde maksimum madde sayısı ile ilgili bir limit olmasa da bu durum sistemin kullanılabilir hafızası ile sınırlandırılmıştır ve teorik olarak 4GB yeterli görülmektedir (Han ve Paek, 2014).

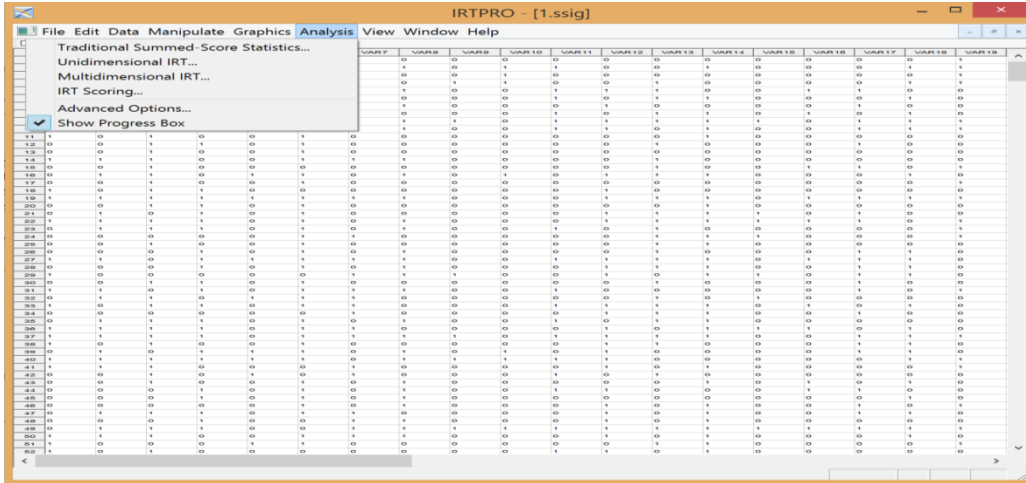
### ***Analizlerin Yapılması***

IRTPRO 2.1 programı bilgisayara yüklendikten sonra IRTPRO ikonu tıkladığında açılan ana menüde "File", "View" ve "Help" seçenekleri bulunmaktadır. Programa "File" sekmesinden elle veri girişi yapılabildiği gibi farklı formatlardaki verilerin aktarımı da sağlanabilmektedir. Farklı formatlı verileri kullanmak için "Import Data" sekmesi seçilir. Bu sekme ile fiks formatlı (.fixed), virgül ile ayrılmış (.csv), boşluk ile ayrılmış (.txt), Excel (.xls) ve SPSS (.sav) gibi dosyalarından veri alınabilir. Program analiz edilecek veriyi, .ssig uzantılı olarak kaydeder. Veri setinde yer alan kayıp veriler de program üzerinden tanımlanabilir. Programın ana ekranı Şekil 1'deki gibidir.

Veri girişi yapıldıktan sonra program, “Analysis” sekmesi altında yer alan (a) geleneksel toplam-puan istatistikleri, (b) tek boyutlu MTK, (c) çok boyutlu MTK, (d) MTK puanlama olmak üzere dört temel analiz türünden seçilene gerçekleştirmektedir. Ayrıca “Analysis” sekmesinde “Advanced Options” ve “Show Progress Box” adlı iki seçenek daha bulunmaktadır (Şekil 2). “Show Progress Box” seçeneği varsayılan olarak işaretlidir. Bu sayede açılan pencere ile programın analiz hangi aşamasında olduğu görülebilir. “Advanced Options” seçeneği analiz penceresindeki “Options” ile aynı içeriktedir ve bu seçenek, ilerleyen kısımda daha ayrıntılı olarak anlatılacaktır.



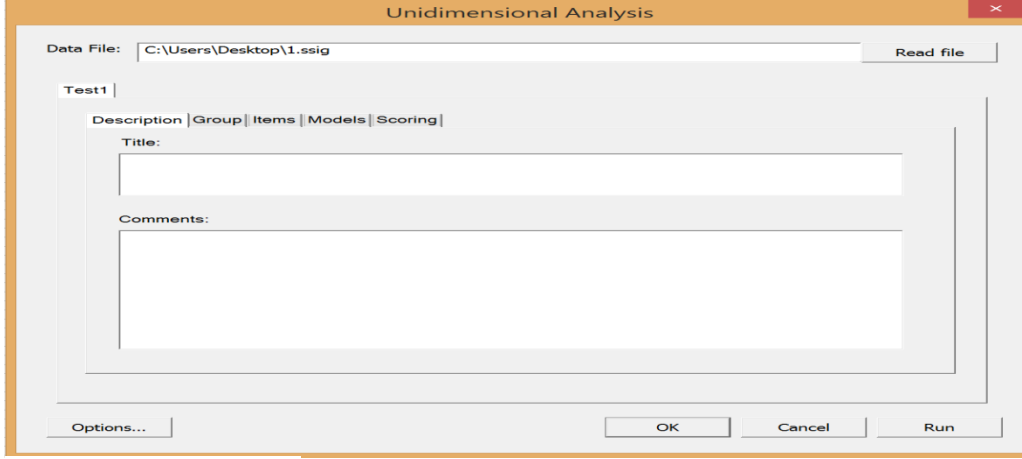
Şekil 1. IRTPRO Ana Ekranı



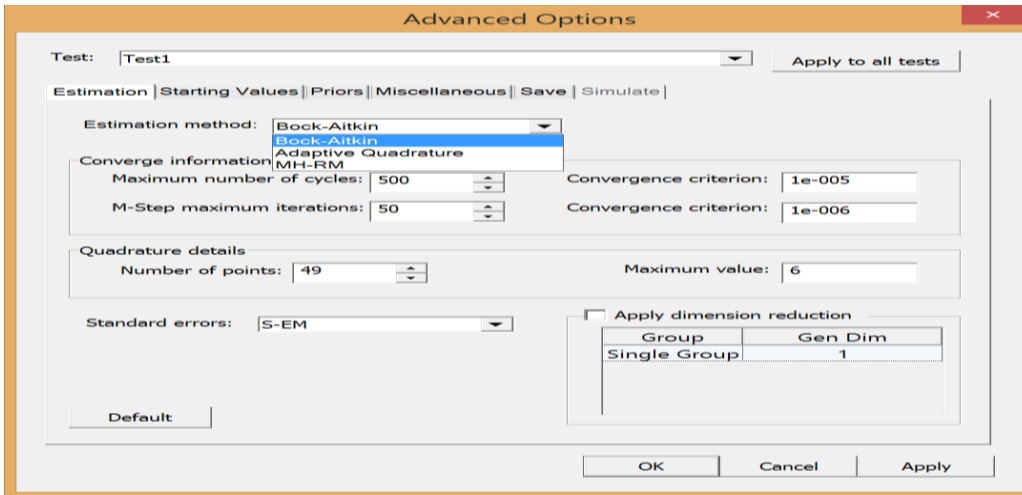
Şekil 2. IRTPRO 2.1’de Analiz Sekmesi

Analiz sekmesinden yapılacak analiz seçildiğinde “Tanımlama” (Description), “Grup” (Group), “Maddeler” (Items), “Modeller” (Models) ve “Puanlama” (Scoring) sekmelerinin yer aldığı pencere açılır (Şekil 3). Bu penceredeki “Tanımlama” kısmından analize isim verilebilir ve analiz ile ilgili yorumlar yazılabilir. Eğer analizler farklı gruplar üzerinden yapılacaksa “Grup” sekmesinden verilerin neye göre gruplandırıldığı ve çoklu grup söz konusu ise referans grubun hangisi olduğu belirtilmelidir. Ancak yapılacak analizde grup değişkeni söz konusu değilse bu sekme atlanabilir. “Maddeler” sekmesinde ise analize hangi maddelerin alınacağı ve çok boyutlu analiz yapıldığı durumda boyut sayısı girilir. “Modeller” sekmesinden 2 parametrelili lojistik (2PL), 3 parametrelili lojistik (3PL), Aşamalı (Graded), Genelleştirilmiş Kısmi Puan (GPCredit), Adlandırılmalı Tepki (Nominal) modellerinden biri seçilir. Eğer analiz edilen veri iki kategorili bir veri ise Aşamalı (Graded), Genelleştirilmiş Kısmi Puan (GPCredit), Adlandırılmalı Tepki (Nominal) seçenekleri kendiliğinden inaktif hale gelecektir. Ayrıca tek boyutlu analizlerde parametre sınırlandırmaları ve değişen madde fonksiyonu için maddelerin tanımlanmasına ilişkin seçimler de buradan yapılabilir. Çok boyutlu analizlerde parametre sınırlandırmaları yoluyla doğrulayıcı faktör analizi, açımlayıcı faktör analizi ve iki-faktör analizlerine ilişkin seçimler de bu sekmeden yapılmaktadır. “Puanlama” sekmesinden yanıt örüntülerini hesaplama yöntemi (EAP ya da MAP), ölçeklemede kullanılacak ortalama, standart sapma, minimum, maksimum, ölçekleme yapılacak dağılım (evren-örneklem)

tanımlanır. Son olarak “Seçenekler” (Options) sekmesi tıklanarak kestirim yöntemleri, başlangıç parametreleri, önseller (prior), kaydedilecek çıktılara ilişkin varsayılan değerler değiştirilebilir (Şekil 4).



Şekil 3. Analiz Penceresi



Şekil 4. İleri Seçenekler Penceresi

Çok boyutlu MTK'ya dayalı veri analizi yapan programlardan TESTFACT'da MAP ve Bayes EAP; NOHARM'da normal ogive modele dayalı olarak en küçük kareler yöntemi, BMIRT'ta MCMC (Markov Chain Monte Carlo), flexMIRT 2'de maksimum marjinal olabilirlik, BA-EM (Bock Aitkin Expectation Maximization) ve MH-RM (Metropolis Hastings Robbins Monro) ile kestirim yapılmaktadır. Bunlara karşılık IRTPRO 2.1 programında madde kalibrasyonu için maximum olabilirlik (Maximum Likelihood-ML) ya da madde parametrelerinin önsel dağılımları tanımlandığı takdirde Maximum a posteriori (MAP) yöntemi kullanılmaktadır. Bu yöntemlerle beraber boyutluluk ve modelin yapısının farklı kombinasyonlarına göre en iyi performansı sağlamak üzere Bock-Aitkin Expectation-Maximization (BA-EM), Adaptive Quadrature (ADQEM) ve Metropolis-Hastings Robbins-Monro (MH-RM) kestirim yöntemleri bulunmaktadır.

Program tarafından otomatik olarak seçili olan Bock-Aitkin EM, tek ve iki boyutlu analizler için kullanılması önerilen bir kestirim yöntemidir. Araştırmacılar, Bock-Aitkin EM'de bulunan quadrature noktalarının sayısını, bu noktaların hangi aralık üzerinde yayıldığını, maksimum döngü sayısını (E-step), maksimum iterasyon sayısını (M-step) değiştirebilirler. Standart hata kestirimi içinse S-EM, M-Step, Xpd, Sandwich olmak üzere dört farklı seçenek bulunmaktadır. Bu kestirim



yöntemi birçok tek boyutlu ve iki faktör MTK modellerinde parametre kestirimi için etkilidir (Cai vd., 2011).

Bock-Aitkin EM algoritmasında bütün bireylerin önsel dağılımları için sabit bir quadrature node olması, kullanılan standart numerik quadraturelar ile ilgili bir problem oluşturmaktadır. Bu durum olabilirlik değerlerini hesaplamada çok sayıda quadrature noktasının kullanımını gerektirir. Bu problemle başa çıkmak için IRTPRO adaptive quadrature denilen sayısal bir integrasyon süreci önermektedir. Bu süreç deneysel Bayes ortalamalarını ve kovaryanslarını kullanmaktadır (Cai vd., 2011). MH-RM algoritması ise boyut sayısı iki ya da üçü aşan çok boyutlu MTK uygulamalarında kullanılmak üzere geliştirilmiştir.

Program ile ölçek puanlarını hesaplamak üzere üç tür MTK kestirim yöntemi önerilmektedir. Buna göre kullanılacak kestirim yöntemleri sırasıyla Bayes kestirimi (Bayes estimation-EAP), toplam puan EAP (Summed Score EAP-SSEAP) ve Bayes model kestirimi (Bayes model estimation-MAP) şeklindedir.

### ***Çıktı Dosyaları, Uyum İyiliği İndeksleri ve Betimsel İstatistikler***

IRTPRO 2.1; geleneksel toplam puana dayalı analizlerde -sss.htm ve MTK'ya dayalı analizlerde -ssc.htm ve -irt.htm uzantılı çıktı dosyaları oluşturmaktadır. -irt.htm uzantısı ile kaydedilmiş bir çıktı dosyasının görünümü şekil 5'teki gibidir. Bu çıktılarda sonuçlar tablolarla ve açıklamalarıyla birlikte yer almaktadır. Bu çıktı dosyalarında kestirim sonuçları, uyum indeksleri gibi bilgiler bulunmaktadır. Ayrıca -irt.irtplot uzantılı dosyada madde karakteristik eğrisi, madde bilgisi, madde karakteristik eğrisi ile madde bilgisinin kombinasyonu, toplam test bilgisi, test karakteristik eğrisi grafikleri yer almaktadır. Birey parametrelerine ilişkin analiz sonuçları ise -sco.txt uzantılı birey parametreleri çıktı dosyası çalışılan klasöre otomatik olarak kaydedilmektedir. Ayrıca ileri seçenekler penceresindeki "Save" seçimi açılarak madde parametre kestirimleri, parametre kestirimlerinin kovaryans matrisi, tek boyutlu modeller için bilgi değerleri, maddeler arası polikorik korelasyonlar, faktör yükleri ve temel çıktılar istenirse ASCII formatında (.txt) çalışılan klasöre kaydedilmektedir. Ayrıca program, bütün komutları otomatik olarak .irtpro dosyasında kaydetmektedir. Kaydedilen bu syntax dosyası üzerinden değişiklikler yapılabileceği gibi syntax dosyası kullanıcı tarafından da oluşturulabilir.

IRTPRO 2.1 yazılımı madde sayısına, cevap kategorilerine ve cevaplara dayalı olarak, -2log olabilirlik (-2log-likelihood), Akaike Bilgi Ölçütü (ABÖ), Bayesian Bilgi Ölçütü (BBÖ) gibi birkaç farklı uyum iyiliği indeksi ile Chen ve Thissen (1997) tarafından geliştirilen yerel bağımlılık istatistiği (local dependence-LD) ve Orlando ile Thissen (2000, 2003) tarafından önerilen madde uyumu istatistikleri ( $S-X^2$ ) gibi betimsel istatistikler hesaplanmaktadır. Yazılım, yukarıdaki indekslere ek olarak Maydeu – Olivares ve Joe (2005) tarafından önerilen ki-kare testi ile ilişkili ve yaklaşık hataların ortalama karekökü değerini (RMSEA) veren  $M_2$  istatistiğini de hesaplamaktadır. Ancak bu indeks, yalnızca BA-EM kullanıldığında analiz çıktılarında yer almaktadır. BA-EM algoritması ile tüm bu uyum indeksleri hesaplanırken, ADQ ve MH-RM algoritmaları kullanıldığında daha az uyum indeksi elde edilmektedir.

Item	Label	a	s.e.	c	s.e.	b	s.e.	logit g	s.e.	g	s.e.	
1	C1	3	0.78	2	2.48	0.23	-3.18	0.94	1	-1.38	0.56	0.20
2	C2	6	0.82	5	0.65	0.21	-0.80	0.35	4	-1.42	0.56	0.19
3	C3	9	1.11	8	-0.21	0.28	0.19	0.23	7	-1.55	0.54	0.17
4	C4	12	0.75	11	0.97	0.19	-1.31	0.44	10	-1.41	0.56	0.20
5	C5	15	0.68	14	1.79	0.19	-2.62	0.81	13	-1.38	0.56	0.20

Şekil 5. IRTPRO 2.1 madde parametreleri çıktı dosyası örneği (-irt.htm)

Şekil 5'te yer alan madde parametreleri çıktısında görüldüğü gibi program ile her bir maddeye ilişkin "a", "c", "b" ve "g" parametreleri kestirilmektedir. IRTPRO 2.1'de "c" parametresi, "b" eşik parametresi ve "a" ayırt edicilik parametresinin etkileşimi ( $b = -c / a$ ) ile elde edilen ve çok boyutlu MTK literatüründe "d" parametresi olarak geçen kesişim (intercept) parametresidir. Eğim-eşik formu çok boyutlu modellerde doğru genellemeler yapamadığı için IRTPRO 2.1'de hem tek boyutlu hem de çok boyutlu bütün modellerde eğim-kesişim formu kullanılmaktadır ( $a(\theta - b) = a\theta + c$ ) (Cai vd., 2011). Ancak tek boyutlu modellerde "b" eşik parametresi de raporlanmaktadır. IRTPRO 2.1 çıktılarında şans parametresinin gösterimi de dikkat çekicidir. Buna göre literatürde "c" parametresi şeklinde gösterilen şans parametresi ise IRTPRO 2.1 çıktılarında "g" parametresi olarak yer almaktadır (Reckase, 2009; Ackerman, 1994; Way, Ansley ve Fosyth, 1988).

### Kestirim Süresi

IRTPRO 2.1 ile yapılan analizlerde kestirilecek parametre sayısına, hesaplanacak istatistiklere, veri setinde yer alan madde/birey sayısına ve veri setinin yapısına bağlı olarak analiz süresi değişiklik göstermektedir. Intel® Core™ i7-4500CPU @1.80GHz 2.40 GHz 8.00GB RAM 64 bit işletim sistemi olan bir kişisel bilgisayarda tek boyutlu ve iki boyutlu yapılarda 20 madde içeren 2000 kişilik veri setleri üretilmiş ve bu veri setlerinin analizi için geçen süre Tablo 2'de verilmiştir.

Tablo 2. IRTPRO 2.1'de Örnek Kestirim Süreleri

Yapı	Model	Geçen Süre (saniye)
Tek boyutlu	1 PLM	0,87
	2 PLM	1,47
	3 PLM	2,85
	Aşamalı Tepki Modeli	4,33
	Genelleştirilmiş Kısmi Kredi M.	4,19
İki boyutlu	1 PLM	19
	2 PLM	56,41
	3 PLM	299,72
	Aşamalı Tepki Modeli	246,99
	Genelleştirilmiş Kısmi Kredi M.	551,16

Tablo 2’de yer alan kestirimlerde işlemci sayısı (Processors) iki olarak ayarlanmıştır ve birey parametreleri için EAP puanlama kullanılmıştır. Diğer seçimlerde varsayılan değerler kullanılmıştır. İki boyutlu kestirimlerde karmaşık yapıli veri setleri kullanılmış ve “Constraints” seçimi tıklanarak “a” parametrelerinde sınırlandırma yapılmıştır. Tablo 2’de yer alan kestirim süreleri incelendiğinde; model sabit kalmak üzere boyut sayısı arttığında birey ve madde parametrelerini kestirmek için gereken sürenin de önemli ölçüde arttığı görülmektedir.

## SONUÇLAR ve TARTIŞMA

IRTPRO 2.1; hem tek boyutlu hem de çok boyutlu MTK’ya dayalı veri analizi yapan programların yerine getirebildiği fonksiyonların tek bir programla yapılabilmesini sağlamaktadır. Program ile iki-faktör ve madde takımı analizleri; hem iki kategorili hem de çok kategorili puanlanan maddelere sahip testlerin analizi de yapılabilir. Yine MTK’ya dayalı olarak değişen madde fonksiyonu, çoklu grup ve çoklu tepki kategorilerinin analizlerinin yapılması mümkündür. Çok çeşitli uygulamaların tek bir programla yapılabilmesi sayesinde aynı veri setinde farklı uygulamalar yapan araştırmacıların kullandıkları veri setini farklı formatlara dönüştürmesine gerek kalmamaktadır. Kullanıcı dostu ara yüzü, araştırmacıların harf ve boşluk duyarlı kod yazmadan sekmelere tıklayarak analiz yapmasını sağlamaktadır. Bu açılarından ele alındığında IRTPRO 2.1, araştırmacıların yazılımı çalıştırmaya ilişkin olarak iş yükünü önemli ölçüde azaltmaktadır.

IRTPRO 2.1’in bu avantajlı yönlerinin yanında programın geliştirilmesine ihtiyaç duyulan yönleri de bulunmaktadır. Örneğin, program ile elde edilen parametre kestirimleri lojistik modele göre yapılmakta ve sonuçlar ölçekleme faktörünü ( $D = 1.7$ ) içermemektedir. Bu sebeple ogive modelde kestirilen parametrelerle IRTPRO 2.1’in kestirdiği parametreleri karşılaştırabilmek için ayırt edicilik parametresi değerlerinin 1.7’ye bölünmesi gerekmektedir ve bu durum farklı program çıktılarıyla doğrudan karşılaştırma yapmanın önünde engel oluşturmaktadır. Program çıktılarına ogive model çıktısı eklenmesinin kullanıcıların için yararlı olacağı düşünülmektedir (Paek ve Han, 2012). Ölçek puanlarının hesaplanması için EAP ve MAP seçenekleri yer almakta; ML yöntemine ilişkin bir seçim bulunmamaktadır. Ayrıca program çıktılarında yer alan kesişim ve şans parametrelerinin simgeleri literatür ile tutarsızlık göstermektedir. İleri seçenekler seçimi ile analizde kullanılacak işlemci sayısı 1 - 8 arasında bir değerde belirlenebilmektedir. Ancak yüksek işlemci sayıları seçildiğinde program, bazı veri setlerinde hata vererek analizi sonlandıramamaktadır. Bu gibi durumlarda düşük işlemci sayısı ile çalışmak problemi ortadan kaldırırsa da bu durum kestirim süresinin uzamasına neden olmaktadır. Programın kullanılan bilgisayarda yönetici modunda çalıştırılması gerektiği için bu durum iş bilgisayarlarında ya da ortak kullanımlı bilgisayarlarda problem oluşturabilmektedir. Son olarak yazılımın alandaki geçmişi fazla değildir ve kullanıcı sayısı sınırlıdır. Araştırmacıların yazılıma ilişkin kaynakları kullanıcı el kitabı ve teknik destekten oluşmaktadır ve farklı tartışmaların, soru cevapların ve uygulama örneklerinin yer aldığı forumlar gibi çok sayıda farklı kullanıcının oluşturduğu kaynakların geliştirilmesine ihtiyaç vardır (Han ve Paek, 2014). Sonuç olarak IRTPRO 2.1’in avantaj ve dezavantajları göz önünde bulundurulduğunda; yazılımın pek çok araştırma sorusunun tek ve çok boyutlu madde tepki kuramına dayalı olarak cevaplanmasında aracı olacağı ve kullanılan yazılıma dayalı olarak ortaya çıkan iş yükünü (veri çekme, kod yazma, çıktıların okunabilirliği) ciddi ölçüde azaltacağı düşünülmektedir. Araştırmacılara bu üstünlükler ve sınırlılıklar açısından yazılımın deneme sürümünü detaylı biçimde incelemeleri ve bundan sonra lisansı satın alıp almamaya karar vermeleri önerilmektedir. İleride yapılacak yazılım inceleme çalışmalarında MTK’ya dayalı analiz yapan yazılımlar ve IRTPRO 2.1 birlikte ele alınarak simülasyon verisi ve gerçek veri üzerinde yapılacak analizlerin çıktıları karşılaştırılabilir, kestirim süresi, uyum indeksleri vs. üzerinden üstünlükleri ve zayıflıkları tartışılabilir.

## KAYNAKÇA

Ackerman, T. A. (1994). Using multidimensional item response theory to understand what items and tests are measuring. *Applied Measurement In Education*, 7(4), 255-278.



- Ansley, T. N., & Forsyth, R. A. (1985). An examination of the characteristics of unidimensional IRT parameter estimates derived from two-dimensional data. *Applied Psychological Measurement*, 9(1), 37-48.
- Bock, R. D., Gibbons, R., Schilling, S. G., Muraki, E., Wilson, D. T., & Wood, R. (2003). *TESTFACT 4.0* [Computer software and manual]. Lincolnwood, IL: Scientific Software International.
- Cai, L. (2013). *flexMIRT Version 2: Flexible multilevel multidimensional item analysis and test scoring* [Computer software]. Chapel Hill, NC: Vector Psychometric Group.
- Cai, L., du Toit, S. H. C., & Thissen, D. (2011). *IRTPRO: Flexible, multidimensional, multiple categorical IRT modeling* [Computer software]. Lincolnwood, IL: Scientific Software International.
- 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.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associate, Inc.
- Fraser, C., & McDonald, R. (1988). NOHARM: Least squares item factor analysis. *Multivariate Behavioral Research*, 23(2), 267-269.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory principles and applications*. Boston-USA: Kluwer-Nijhoff Publishing.
- Han, K. T., & Paek, I. (2014). A review of commercial software packages for multidimensional IRT modeling. *Applied Psychological Measurement*, 38(6), 486-498.
- Linacre, J. M. (2011). *Winsteps Rasch measurement* [Computer program]. [Çevrim-içi: <http://www.winsteps.com/index.htm> ], Erişim tarihi: 26 Nisan 2015.
- Maydeu-Olivares, A., & Joe, H. (2005). Limited and full information estimation and testing in 2<sup>n</sup> contingency tables: A unified framework. *Journal of the American Statistical Association*, 100, 1009–1020.
- Muraki, E., & Bock, R. D. (2003). *PARSCALE 4: IRT item analysis and test scoring for rating-scale data* [Computer software]. Chicago, IL: Scientific Software International.
- Orlando, M., & Thissen, D. (2000). Likelihood-based item fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, 24(1), 50-64.
- Orlando, M., & Thissen, D. (2003). Further investigation of the performance of S-X2: An item fit index for use with dichotomous item response theory models. *Applied Psychological Measurement*, 27(4), 289-298.
- Paek, I., & Han, K.T. (2012). IRTPRO 2.1 for windows (Item response theory for patient-reported outcomes). *Applied Psychological Measurement*, 37(3), 242-252.
- Reckase, M. D. (2009). *Multidimensional item response theory (Statistics for social and behavioral sciences)*. New York: Springer.
- Rizopoulos, D. (2015). *Package "ltm", Latent Trait Models under IRT*. [Çevrim-içi: <https://cran.r-project.org/web/packages/ltm/ltm.pdf>], Erişim tarihi: 29 Nisan 2015.
- Rusch, T., Mair, P., & Hatzinger, R. (2013). *Psychometrics with R: A review of CRAN packages for item response theory*. [Çevrim-içi: <http://epub.wu.ac.at/4010/>], Erişim tarihi: 10 Mayıs 2015.
- Thissen, D. (2003). *MULTILOG 7: Multiple categorical item analysis and test scoring using item response theory* [Computer software]. Chicago, IL: Scientific Software International.
- Way, W. D., Ansley, T. N., & Forsyth, R. A. (1988). Unidimensional IRT estimates the comparative effects of compensatory and noncompensatory two-dimensional data on unidimensional IRT estimates. *Applied Psychological Measurement*, 12(3), 239-252.
- Yao, L. (2003). *BMIRT: Bayesian multivariate item response theory* [Computer software]. Monterey, CA: CTB/McGraw-Hill.
- Zimowski, M. F., Muraki, E., Mislavy, R. J., & Bock, R. D. (2003). *BILOG-MG: Multiple-group IRT analysis and test maintenance for binary items (Version 3)* [Computer software]. Chicago, IL: Scientific Software International.

## EXTENDED ABSTRACT

### Introduction

From the initial introduction to the end of 1970s, the Item Response Theory (IRT) used for unidimensional tests was extended to multi-dimensional tests from late 1970s and early 1980s and applied to tests measuring multiple abilities under Multidimensional Item Response Theory (MIRT) (Ansley & Forsyth, 1985; Reckase, 2009). Differently from unidimensional IRT, item and ability

parameters are estimated using MIRT equalities in multidimensional tests. The new theory therefore required to develop and use new software.

IRTPRO 2.1 (Item Response Theory for Patient-Reported Outcomes) program was developed by Li Cai, David Thissen and Stephen du Toit in 2011. IRTPRO 2.1 is a statistical software that uses item response theory for item calibration and test scoring. IRTPRO 2.1 uses IRT models that are specified below (Cai, Thissen & du Toit, 2011):

- Two-parameter logistic model (2PL) [becomes one-parameter logistic model with use of equal discrimination parameter (1PL)]
- Three-parameter logistic model (3PL)
- Graded Response Model
- Generalized Partial Scoring Model
- Nominal Response Model

Both unidimensional and multidimensional tests which belong to these models can be analyzed with IRTPRO 2.1. Also a test may also include different combinations of these models depending on the objective of the survey and the nature of measuring instrument.

### ***Program Installation, Analyses, Outputs and Estimation Time***

IRTPRO 2.1 can be used with Windows7, Vista and XP operating systems. Program can be accessed at <http://www.ssicentral.com>. The trial version of the software is also accessed at the same address with no cost. Users who use the paid version of the program are provided with technical support.

Data can be manually entered in IRTPRO 2.1 program via “File” tab as well as data can be extracted from files such as Excel (.xls) and SPSS (.sav) with fixed format (.fixed), comma separated (.csv) and space separated (.txt). Program saves the data that is extracted and analyzed with .ssig extension.

After data is entered, program executes the analysis selected from four basic types included under “Analysis” tab; (a) traditional summed-score statistics, (b) unidimensional IRT, (c) multidimensional IRT, (d) IRT scoring.

In IRTPRO 2.1 program, if maximum likelihood (ML) or prior distribution of item parameters is defined for item calibration, maximum a posteriori (MAP) method is used. With these methods, Bock-Aitkin Expectation-Maximization (BA-EM), Adaptive Quadrature (ADQEM) and Metropolis-Hastings Robbins-Monro (MH-RM) estimation methods are available to achieve the best performance based on the dimensionality and the different combinations of model structure.

Three types of IRT estimation methods are suggested with program to calculate scale scoring. So, the available estimation methods are Bayes estimation (EAP), Summed Score EAP (SSEAP) and Bayes model estimation (MAP), respectively.

IRTPRO 2.1 produces output files with -sss.htm extension in traditional summed score based analyses as well as output files with -ssc.htm and -irt.htm extension in analyses based on IRT. In addition, the file with -irt.irtplot extension contains item and test information curves. In analysis of individual parameters, individual parameters with -sco.txt extension will be automatically saved in the folder running the output files. Also, if “Save” option is opened in the next options window, the chosen outputs will be saved in the running folder as ASCII format (.txt).

Based on the number of items, response categories and responses, IRTPRO 2.1 software computes several different types of goodness of fit index such as -2log-likelihood, Akaike's Information

Criterion (AIC), Bayesian Information Criterion (BIC), and descriptive statistics such as local dependence-LD developed by Chen and Thissen (1997) and S-X2item compatibility statistics. In addition to indexes above, the software also computes the M2 statistics that provides chi-square test and root mean square error of approximation (RMSEA). However, this index is included analysis outputs only when BA-EM is used. While all these fit indexes are calculated with BA-EM algorithm, less fit index is achieved when ADQ and MH-RM algorithms are used.

In analyses performed with IRTPRO 2.1, the analysis time varies depending on the number of parameters to be estimated, statistics to be calculated, the number of items/individuals in the data set, and the structure of data. Simulated uni- and two-dimensional data sets with 20 items and 2000 examinees analyzed with a personal computer with Intel® Core™ i7-4500CPU @1.80GHz 2.40 GHz 8.00GB RAM 64 byte operating system. Table 1 presents the time for analyzing these data sets.

The number of processors was set at 2 for the estimations in Table 1, and EAP scoring was used for individual parameters. Default values were used for other options. Complex structured data sets were used for two-dimensional estimations and “a” parameters were constrained by clicking on the “Constraints” option. As can be seen in Table 2, when model remained constant, the time needed to estimate ability and item parameters was significantly increased as the number of dimensions was increased.

Table 1. Example Estimation Times in IRTPRO 2.1

Structure	Model	Estimation Time (second)
One-dimensional	1 PLM	0,87
	2 PLM	1,47
	3 PLM	2,85
	Graded Response Model	4,33
	Generalized Partial Credit Model	4,19
Two-dimensional	1 PLM	19
	2 PLM	56,41
	3 PLM	299,72
	Graded Response Model	246,99
	Generalized Partial Credit Model	551,16

### ***Results and Discussion***

IRTPRO 2.1 allows executing functions, which can be performed by several programs for data analysis based on IRT, with only one software. User-friendly interface allows researchers to perform analyses by clicking on tabs without typing case-sensitive and space-sensitive codes.

In addition to such advantageous of IRTPRO 2.1, it has aspects that require developing the program. For example, parameters obtained from the program are estimated based on logistic model and do not include scaling factor ( $D = 1.7$ ). There is no option available for ML method to calculate scale scores. Furthermore, some parameter terms on the program outputs conflict with representations in the literature. When high number of processors are selected, program cannot terminate the analysis and gives errors in some data sets. Finally, the program can be run only on administrator mode.

As a result, when advantages and disadvantages of IRTPRO 2.1 considered; it can be expressed that the software appears to help answering many survey questions based on unidimensional and multidimensional item response theory and can reduce the problems due to the software which is being used. In addition, researchers are recommended to buy the software license after reviewing the trial version in detail.