

Test Boyutluluğunun Analizinde Kullanılan Programların İncelenmesi

An Overview of Software for Assessing Test Dimensionality

Güler YAVUZ * Nuri DOĞAN**

Öz

Son yıllarda çok boyutlu veri setlerine ve çok boyutlu veri setleriyle yapılan çeşitli test uygulamalarına (test eşitleme, alt test puanlama, değişen madde fonksiyonu gibi) olan ilgi önemli düzeydedir. Bu çalışmada test boyutluluğunun analizinde ve çok boyutlu veri setleriyle çeşitli test uygulamalarında yaygın kullanılan bazı popüler paket programları (Mplus, NOHARM, TESTFACT, IRTPRO, flexMIRT, BMIRT, DIMTEST, DETECT, CCPROX/HAC, Velicer' in MAP testi ve Paralel Analiz) incelenmiştir. Bu programların ulaşılabilirliği, hangi tür veri setleri, modeller, kestirim teknikleri ile analiz yapabildikleri belirlenmiştir. Programların boyutluluk analizini hangi tür yaklaşımla (açımlayıcı, doğrulayıcı) yaptıkları ve program çıktılarında bulunan indekslerin nasıl yorumlandıkları incelenmiştir. Ayrıca programların çeşitli özellikleri açısından karşılaştırılmalarına yönelik yapılmış çalışmalara yer verilmiştir.

Anahtar Kelimeler: boyutluluk, Mplus, IRTPRO, flexMIRT, BMIRT

Abstract

In recent years, there is substantial demand for multidimensional data sets and their various test applications (such as equating test scores, subtest score reporting, differential item functioning). In this study, some popular software packages (Mplus, NOHARM, TESTFACT, IRTPRO, flexMIRT, BMIRT, DIMTEST, DETECT, CCPROX/HAC, Velicer's MAP test and Parallel Analysis) which are used for analyzing test dimensionality and used for various test applications with multidimensional data sets, was investigated. The software packages was investigated in terms of their availability, their capabilities of data sets, models, estimation techniques. The software packages' dimensionality analyses approach (confirmatory, exploratory) was investigated. Also assessing of the indices which can be gained from software packages' outputs was investigated. Finally the simulation studies in which the packages are compared in terms of their various properties was reviewed.

Key Words: dimensionality, Mplus, IRTPRO, flexMIRT, BMIRT

GİRİŞ

Testlerin veya testlerde bulunan maddelerin sadece tek bir örtük özelliği ölçtüğünü varsaymak ve bu varsayım altında ölçmeler yaparak, bireyler hakkında kararlar vermek 1980' lerden sonra daha da tartışılır hale gelmiştir (Ackerman, 1989; Ansley ve Forsyth, 1985; Drasgow ve Parsons, 1983; Harrison, 1986; Way, Ansley ve Forsyth, 1986). Özellikle 1990' lardan itibaren çok sayıda araştırmacı (Sireci, Thissen, ve Wainer, 1991; Yen, 1993; Bradlow, Wainer, ve Wang, 1999; Chen ve Thissen, 1997; De Champlain, 1996; Shealy ve Stout, 1993; Walker ve Beretvas, 2003) çok boyutlu veri setlerini tek boyutlu model uygulamalarıyla incelemiştir. Bu çalışmaların büyük bir çoğunluğundan çıkan sonuçlara dayanarak, araştırmacılar çok boyutlu madde tepki kuramı model, program ve uygulamalarının geliştirilmesi gerektiğini belirtmişlerdir (Ackerman, 1996; Adams, Wilson ve Wang, 1997; Embretson, 1997; McDonald, 1997; Reckase, 1997; Beguin & Glas, 2001; Bolt ve Lall, 2003; Ackerman, Gierl ve Walker, 2003; Walker ve Beretvas, 2003; Yao ve Boughton, 2009; Reckase, 2009).

* Yrd. Doç.Dr., Adıyaman Üniversitesi, Eğitim Fakültesi, Adıyaman-Türkiye, e-posta: gyavuz2010@gmail.com

**Doç.Dr., Hacettepe Üniversitesi, Eğitim Bilimleri Bölümü, Ankara-Türkiye, e-posta: nurid@hacettepe.edu.tr

Özellikle son yıllarda, test uzmanlarının ve test merkezlerinin odağında başarı ve yetenek testlerine ait çok boyutlu veri setleri ile dikey ölçekleme, test eşitleme, alt test puanlama gibi temel uygulamalara ilişkin daha doğru analizler yapma çabası bulunmaktadır. Test uygulamalarıyla elde edilen sonuçların doğruluğu kullanılan programlarla yakından ilişkilidir. Araştırmacılar ve uygulayıcılar, program seçimini test koşullarına uygun yapmak durumundadır. Farklı modellerle ve test koşulları ile analiz yapabilen programların sahip oldukları kestirim teknikleri sonuçların doğruluğu üzerinde özellikle belirleyicidir.

Bu durum, verilerin boyutluluğunu belirleyebilen, çok boyutlu veri setleriyle çeşitli test uygulamaları yapma olanağı tanıyan, madde ve yetenek parametrelerine yönelik farklı kestirim algoritmalarına ve yollarına sahip bilgisayar programlarının geliştirilmesine yönelik de önemli bir araştırma alanı ortaya çıkarmaktadır.

Günümüzde hızla gelişen bilgisayar teknolojisinden dolayı ve her gün daha iyi ve yeni bir program ve kestirim yöntemi elde etmek mümkün olabildiğinden kalıcı bir program listesi yapmak kolay değildir. Bilgisayar programlarına ilişkin listenin sürekli güncellenmesi gerekmektedir. Bu sebeple çeşitli amaçlarla çok boyutlu modellerle ve veri setleriyle analiz yapmak isteyen araştırmacıların iyi bir alanyazın araştırması yapması gerekmektedir. Boyutluluk analizinde kullanılabilen istatistik programları arasında önemli farklılıklar bulunmaktadır. Herhangi bir araştırmacı bu programlardan biriyle çalışmadan önce programların kullandığı kestirim tekniği, kestirim süresi, yazılım dili, maliyet, ne tür veri setleriyle analiz yapılabildiği, hangi modellerin kullanılabildiği vb. gibi çok sayıda avantaj ve dezavantajları bilmek durumundadır.

Araştırmanın Amacı

Bu çalışmanın amacı, çeşitli disiplinlerle çalışan araştırmacılara ve uygulayıcılara boyutluluğun değerlendirilmesinde kullanılabilen mevcut paket programların ve yöntemlerin tanıtılmasıdır. Bu araştırmada ilk olarak programa erişilebilir yolları, programın kullanımına yönelik dökümanların var olup olmadığı, varsa neler olduğu ele alınmış, ardından programların teknik özellikleri incelenmiştir. Çok boyutlu veri setlerinin boyutluluk analizini sağlayan ve bu veri setleriyle çeşitli test uygulamaları yapma olanağı tanıyan programların hangi tür verilerle (ikili, çok kategorili), modellerle, kestirim teknikleri ve algoritmalarıyla ne tür analizler yapabildikleri incelenmiştir. Bunun yanı sıra programların çıktılarında yer alan indekslerin hangileri oldukları ve bu indekslerin yorumlanmasına ilişkin ölçütler açıklanmıştır. Son olarak programların çeşitli uygulamalar açısından avantaj ve dezavantajlarını inceleyen, programları çeşitli özellikleri açısından karşılaştıran simülasyon çalışmaları verilmiştir.

Araştırmada sırasıyla Mplus, NOHARM, TESTFACT, IRTPRO, flexMIRT, BMIRT, DIMPACK (DETECT, DIMTEST, CCPROX/HCA), Velicer'in MAP Testi, Paralel Analiz olmak üzere boyutluluk analizinde yaygın kullanılan programlar ve yaklaşımlar incelenmiştir. Ayrıca programlara ilişkin özelliklerin karşılaştırıldığı bir tablo Ek 1'de verilmiştir.

PROGRAMLAR

Mplus 7.31:

Mplus programı Muthen ve Muthen (1998-2012) tarafından geliştirilmiştir. Programın 1998-2014 yılları arasında farklı sürümleri geliştirilmiş olup, 7.31 sürümü 2014'de geliştirilmiş ve 7.4 sürümünün 2015 yılının sonlarında kullanıcılara sunulması hedeflenmektedir. Programın kullanımına yönelik olarak ayrıntılı yazılmış bir pdf dosyası bulunmaktadır. Mplus ücretli olup, hem nasıl edinilebileceğine ilişkin ayrıntılar hem de kullanımına ilişkin pdf dosyası geliştiriciler tarafından hazırlanan bir web sitesinde bulunmaktadır.

(<https://www.statmodel.com/ugexcerpts.shtml>).

Mplus, araştırmacılara çok sayıda model seçeneği, kestirim algoritması, grafik seçenekleri ve verileri analiz etme olanağı sağlayan bir istatistiksel modelleme programıdır. Program tek ve çok düzeyli verilerle, kayıp verilerle, sürekli, sıralı, sınıflama ve oran ölçeğiyle elde edilmiş veri setleriyle ve tüm bu veri türlerinin kombinasyonlarıyla analiz yapma olanağı vermektedir. Ayrıca Mplus programı ile veriyi Monte Carlo simülasyonları kullanarak üretmek de mümkündür (Muthen ve Muthen, 1998-2012). Mplus programı sürekli örtük değişkenler kullanarak regresyon analizi, path analizi, açımlayıcı ve doğrulayıcı faktör analizi, madde tepki kuramı modellemeleri, yapısal eşitlik modellemeleri, örtük büyüme modellemeleri, kesikli zaman serisi ve sürekli zaman serisi analizleri yapmaya olanak vermektedir. Aynı zamanda kategorik örtük değişkenler kullanarak regresyon analizi, path analizi, örtük sınıf analizi, doğrulayıcı örtük sınıf analizi, loglinear modellemesi, çoklu grup modelleme, CACE (Complier Average Casual Effect) modelleme, kesikli ve sürekli zaman serileri analizleri yapmaya da olanak vermektedir. Dolayısıyla hem sürekli hem de kategorik örtük değişkenleri birlikte kullanarak rastgele etkilerle örtük sınıf analizleri, faktör karışımı modellemesi, yapısal eşitlik karışım modellemesi, büyüme karışım modellemesi, kesikli ve sürekli zaman serisi karışım modellemesi yapılabilmektedir (Muthen ve Muthen 1998-2012).

Mplus programı hem Bayesyen hem de olabilirlik (likelihood) yaklaşımlarını kullanarak kestirimde bulunmaktadır. Bayesyen yaklaşımda MCMC (Monte Carlo Markov Chain) algoritmasını, olabilirlik yaklaşımda ise maksimum olabilirlik ve ağırlıklandırılmış en küçük kareler algoritmalarından seçileni kullanmaktadır. Ayrıca Quasi-Newton, Fisher scoring, Newton-Raphson, Expectation Maximization (EM) gibi optimizasyon algoritmaları da seçilebilmektedir (Gelman, Meng ve Stern, 1996; Svetina ve Levy, 2012)

Mplus programı açımlayıcı ve doğrulayıcı yaklaşımlarla boyutluluk analizi yapabilmektedir. Açımlayıcı faktör analiziyle, dik ve eğik döndürme teknikleri uygulanabilmekte, program çıktılarında veri yapısına uygun (polikorik, tetrakorik vb.) korelasyon matrisleri, artık korelasyon matrisleri için özdeğerler verilmektedir. Ayrıca boyutluluğun değerlendirilmesi için RMSR (root mean squared residual), χ^2 istatistiği ve RMSEA (root mean square error of approximation) değerleri programın çıktılarında verilmektedir (Muthen ve Muthen, 1998-2012).

Boyutluluk analizlerinde kullanılan uyum iyiliği ve hata (uyumsuzluk) indekslerinin yorumlanmasına ilişkin farklı görüşler bulunmakla birlikte, çoğunlukla iyiliği indekslerinin değeri 1'e yaklaştıkça uyumun arttığı, diğer bir deyişle modelin veriye uyumunun arttığı, hata indekslerinin ise sifra yaklaştıkça model uyumunun arttığı ifade edilmektedir. χ^2 (Ki-kare) uyum istatistiği örneklem büyüklüğüne duyarlılığı nedeniyle (Bentler ve Bonnet, 1980; Jöreskog ve Sörbom, 1993; Kenny ve McCoach, 2003) tartışılabilir da yaygın kullanılan uyum istatistiklerinden biridir. Örneklem büyüklüğüne duyarlılığı azaltmak için Wheaton Muthen, Alwin ve Summers (1977) tarafından χ^2 / df oranının kullanılması gerektiği önerilmiştir ve bu oran 2' den küçük değerlere sahipse modelin iyi uyum gösterdiği, 2 ile 5 arasında ise modelin kabul edilebilir olduğu ifade edilmektedir (Tabachnick ve Fidell, 2007).

RMSEA uyum istatistiği Steiger (1990) tarafından geliştirilmiş ve bu indeks de kestirilen parametre sayısına duyarlı olmasına rağmen son yıllarda model uyumunu değerlendirme de yaygın kullanılan indekslerden biridir (Diamantopoulos and Sigauw, 2000). RMSEA değerinin 0.10'un üstünde olması kötü uyumu, 0.08 ile 0.10 arasında olması kabul edilebilir uyumu (MacCallum, Browne ve Sugawara, 1996) göstermektedir. Ancak son yıllarda RMSEA değerinin 0.00 ile 0.06 veya 0.07 arasında olması gerektiği ifade edilmektedir (Hu ve Bentler, 1999; Steiger, 2007). RMR ve SRMR değerlerinin 0 olması mükemmel uyumu gösterirken 0.05- 0.08 aralığında ki değerler kabul edilebilirdir (Hu ve Bentler, 1999 Byrne, 1998; Diamantopoulos ve Sigauw, 2000). GFI, AGFI, CFI gibi indekslerin ise 0.90 üstü olması kabul edilebilir bir uyum iyiliği değerini, 0.95 üstü olması ise iyi bir uyum iyiliğinin göstergesi olarak kabul edilir (Tabachnick ve Fidell, 2007). AIC (Akaike Bilgi Kriteri) ve BIC (Bayesyen Bilgi Kriteri) indekslerinin ise daha küçük değerler olması iyi uyumu göstermektedir.

NOHARM(Normal Ogive Harmonic Analysis Robust Method)

NOHARM (Normal Ogive Harmonic Analysis Robust Method) bir model olarak McDonald (1967) tarafından geliştirilmiş olup, Fraser (1998) tarafından bilgisayar programı (<http://noharm.software.informer.com/download/>) olarak geliştirilmiştir. Modele yönelik ayrıntılı bilgilere McDonald (1962; 1967; 1981; 1997; 2000) tarafından yapılan çalışmalarda erişebilmek mümkündür. Program iki kategorili verilerle tek ve çok boyutlu normal ogive modeller için analiz yapabilmektedir. Ancak çok kategorili ve kayıp verilerle analiz yapamamaktadır. Program en küçük kareler kestirimi algoritmasını kullanmaktadır Program ücretsizdir. Programın temel dezavantajlarından biri şans parametresi kestirimi yapamamasıdır ancak kullanıcıların şans parametresi değerlerini programa manuel olarak girmesine imkan tanımaktadır. (McDonald ve Fraser, 1988; Reckase, 2009; Svetina ve Levy, 2012).

Program faktör çözümleri için dik ve eğik döndürme olanağı tanıyarak açımlayıcı ve doğrulayıcı faktör analizi yapma olanağı tanımaktadır. NOHARM tetrakorik korelasyonları kullanmamakta, normal ogive modellere polinomiyal bir yaklaşım geliştirmektedir ve ağırlıklandırılmış en küçük kareler yaklaşımıyla kestirim yapmaktadır (Çok boyutlu modellere polinomiyal yaklaşım için Reckase (2009) incelenebilir). Program çıktılarında beklenen ve gözlenen oranlar arasındaki farka ilişkin artık matrislerini, RMSR (root mean square residual), Tanaka'nın uyum iyiliği indisi (GFI), döndürülmüş faktör yükleri verilmektedir. Ayrıca NOHARM programının çıktılarındaki değerler kullanarak $X^2_{G/D}$ istatistiği ile ALR (approximate likelihood ratio) istatistiği de hesaplanmaktadır. NOHARM çıktılarında yedi bölüm bulunmaktadır. İlk bölümde madde sayısı, analiz başlığı korelasyon matrisleri, kullanıldıysa sabitlenmiş şans parametreleri, bulunmaktadır. İkinci bölümde örtük yetenek ve genel faktör parametreleri bulunmaktadır. Bu parametreler; güçlük ve ayırt edicilik parametreleri ile RMSR ve Tanaka'nın ağırlıklandırılmamış en küçük kareler uyum iyiliği indisidir.

NOHARM programı k boyutlu normal ogive modeli aşağıda verilen eşitlikteki gibi kullanır.

$$P\{U_i = 1 | \Theta = (\theta_1, \dots, \theta_k)\} = N\{\beta_{i0} + \beta_{i1}\theta_1 + \dots + \beta_{ik}\theta_k\}$$

Verilen eşitlikte N normal ogive fonksiyonunu, θ örtük yetenek vektörünü, β_{i0} madde güçlük parametresini, β_{ik} madde ayırt edicilik parametresini göstermektedir.

NOHARM programı ile doğrulayıcı yaklaşımla boyutluluk analizi yapılmışsa model uyumunu değerlendirmek için RMSR (root mean square residual) ve Tanaka'nın uyum iyiliği incelenebilir. RMSR değerinin sıfır olması mükemmel uyumu gösterirken, değer büyüdükçe model uyumu kötü hale gelmektedir (Kline, 2005). Tanaka'nın indeksi 0 ile 1 arasında değerler almaktadır. Bu değer 1'e yaklaştıkça modelin uyumu artmaktadır. Boyut sayısına açımlayıcı yaklaşımla karar verme sürecinde uyum indeksinde hızlı artış gözlenirse ve RMSR değerindeki düşüşler %10 veya daha fazla ise en uygun modelin belirlendiği görüşü bulunmaktadır (Tate, 2003). Finch ve Habing (2005), 1 ve M model arasında bir seri modelin uyumunu belirleyerek, farklı sayıda boyut (örtük değişken) öneren modeller için $X^2_{G/D}$ farkının belirlenmesini ve yorumlanmasını önermişlerdir.

TESTFACT 4.0

TESTFACT modeli ilk olarak "full information item factor analysis" çalışmalarıyla Bock, Gibbons ve Muraki (1998) tarafından oluşturulmuştur. TESTFACT 4.0 sürümü ise Bock, Gibbons, Schilling, Muraki, Wilson ve Wood tarafından (2003) geliştirilmiştir. Program ücretlidir ve araştırmacılar tarafından edinilebilmesi için TESTFACT programının yanı sıra çok sayıda farklı istatistik programı da geliştiren Uluslararası Bilimsel Paket programı (Scientific Software International, <http://www.ssicentral.com/irt/index.html>) grubunun internet sitesinin incelenmesi gerekmektedir. Programın kullanımına yönelik ayrıntılara Mathide Du Toit (2003)'in editörlüğünü yaptığı "IRT

from SSI: BILOG-MG, MULTLOG, PARSCALE, TESTFACT” isimli kitapta Bock ve Schilling (2003) tarafından yazılan “IRT based item factor analysis” isimli bölümde yer verilmiştir.

TESTFACT programı iki ve çok kategorili veriler, kayıp veriye sahip veri setleri tetrakorik korelasyon matrisleri için analiz yapabilmektedir. Program kestirim algoritması olarak Bock ve Aitkin (1981) tarafından geliştirilen marjinal maksimum olabilirlik kestirimini ve ayrıca yetenek parametreleri için Bayesyen algoritması kullanılmaktadır. TESTFACT programının sınırlılıklarından biri doğrulayıcı yaklaşımla boyutluluk analizinde sadece iki faktörlü model için çözüm üretmesi, diğeri ise NOHARM programı gibi şans parametresi kestirimi yapamamasıdır. Ancak TESTFACT programına da şans parametresi için sabit değerlerin kullanıcı tarafından girilebilmesi mümkündür. Program klasik madde analizi, test puanı hesaplama ve maddeler arası tetrakorik korelasyonu kullanarak faktör analizi yapabilmektedir. Ayrıca program ile madde tepki kuramına dayanan modern faktör analizi yapmak mümkündür (Reckase, 2009; du Toit, 2003; Svetina ve Levy, 2012).

TESTFACT programı açımlayıcı faktör analiziyle dik ve eğik döndürme yapmaya olanak tanımaktadır. Boyutluluğun analizinde modele bir faktör daha eklendiğinde Ki-kare (χ^2) uyum istatistiğinde meydana gelen değişiklik temel alınmaktadır. Analizlere tek faktörlü modelle başlanmakta ve modele bir faktör daha eklendiğinde model uyumundaki düzeltilmeler incelenmektedir. Modele bir faktör daha eklendiğinde model uyumunda önemli bir değişiklik olmaz ise yapının tek boyutlu olduğu ifade edilmektedir. Modele bir faktör daha eklendiğinde model uyumunda anlamlı değişiklik olursa testin çok boyutlu olduğuna karar verilmekte ve faktör eklenmeye devam edilmektedir. Testin boyut sayısı eklenen faktörün anlamlı düzeyde değişiklik meydana getirmediği durum olarak kabul edilmektedir.

IRTPRO 2.1

IRTPRO programı Cai, du Toit ve Thissen (2011) tarafından geliştirilmiştir. Programın kullanımına ilişkin ayrıntılı bir pdf dosyası yazarlar tarafından hazırlanmıştır. Program ücretlidir ve araştırmacılar tarafından programın edinilebilmesi ve pdf dosyasının indirebilmesi için Uluslararası Bilimsel Paket programı (Scientific Software International, <http://www.ssicentral.com/irt/index.html>) grubunun internet sitesinin incelenmesi gerekmektedir. Programın kullanımına yönelik ayrıntılı bir pdf dosyası ise internet sitesindeki (<http://coeweb.gsu.edu/coshima/EPRS8410/IRTPRO%20Guide.pdf>) linkinde bulunmaktadır. Program iki ve çok kategorili verilerle, erişilmemiş veya kayıp veriye sahip veri setleriyle, tek ve çok boyutlu madde tepki kuramı modellerinin birçoğuyla analizler yapmaya olanak tanımaktadır.

IRTPRO programına çeşitli istatistik paket programlarından veri aktarılabildiği gibi fixed, csv, txt ve xls uzantılı formattaki verilerde aktarılabilmekte ve IRTPRO çıktısı dosyalar .ssig* uzantılı olarak kaydedilmektedir. IRTPRO açımlayıcı ve doğrulayıcı analizleri tek ve çok boyutlu modellerin birçoğuyla (bir, iki ve üç parametrelili lojistik modeller, aşamalı tepki modeli, genelleştirilmiş kısmi puan modeli, sınıflamalı model) yapabilmektedir. IRTPRO programı parametre kestiriminde genellikle maksimum olabilirlik tekniğini kullanılmaktadır ancak madde parametreleri için önsel dağılımların önceden belirlenmesi durumunda maksimum sonsal (Maximum a posteriori) kestirim tekniği kullanılabilir. Ayrıca her biri farklı test yapılarıyla veya modellerle daha iyi sonuçlar veren çeşitli kestirim teknikleri (Bock Aitkin EM (BA-EM), Bifactor EM, Generalized Dimension Reduction EM, Metropolis Hasting Robbins-Monro (MH-RM)) ile IRTPRO programında madde parametresi kestirimi olanaklıdır.

Madde sayısı, cevap kategorileri ve cevaplayıcılara göre değişmek suretiyle çeşitli uyum iyiliği istatistikleri madde kalibrasyonundan sonra hesaplanmaktadır. 2loglikelihood, Akaike Bilgi Kriteri (Akaike Information Criterion (AIC)) ve Bayesyen bilgi kriteri (BIC) rapor edilen başlıca istatistiklerdir. Bazı modeller için ayrıca M_2 istatistiği hesaplanmaktadır. Bilişsel istatistikler ise LD (yerel bağımsızlık istatistiği) ve $SS-X^2$ madde uyum istatistikleridir.

flexMIRT 2.0

flexMIRT 2.0 programı Cai (2013) tarafından geliştirilmiştir. Program çok zengin psikometrik ve istatistik özelliklere sahiptir ve ücretlidir. Programın özellikleri ve kullanımına yönelik ayrıntılı bir pdf dosyası Houts ve Cai (2013) tarafından yazılmıştır. Programın tüm özelliklerine yönelik ayrıntılı bilginin olduğu ve programın satın alınabileceği bir web sitesi bulunmaktadır (<https://www.vpgcentral.com/irt-software/>). Programın kullanımına hızla teknik destek sunan bir internet sitesi de bulunmaktadır (<http://www.vpgcentral.com/irt-software/support/>).

Program Windows temellidir, grafik ara yüzü sayesinde kullanıcı dostudur. Oldukça hızlıdır ve boşluk, virgül ve sekmeyle ayrılmış veriyi yüklemeye olanak sağlamakta, iki ve çok kategorili verilerle analiz yapabilmektedir. Monte Carlo simülasyonu ile çok sayıda modele ilişkin veri üretmek de mümkündür. Ayrıca program kayıp ve erişilmemiş verisi olan veri setleriyle de analiz yapabilmektedir. flexMIRT programı doğrulayıcı yaklaşımla çok sayıda tek ve çok boyutlu MTK modeli için madde parametresi kestirimi yapabilmektedir (Örneğin, tek boyutlu iki kategorili modellerden Rasch, bir, iki ve üç parametrelili lojistik model, Samejima'nın aşamalı tepki modeli, sınıflama modeli, kısmi ve genelleştirilmiş kısmi puan modeli ile dereceli ölçek modeli, çok boyutlu bir, iki, üç parametrelili lojistik model, çok boyutlu aşamalı tepki modeli gibi). Ayrıca tek ve çok gruplu, tek ve çok düzeyli veri setlerine ilişkin parametre kestirimi de yapılabilmektedir. Madde parametresi kestirimi için BA-EM ve MHRM tekniklerini kullanılarak flexMIRT programı ile parametre kestirimi yapılabilmektedir. Yetenek parametresi kestirimi için maksimum olabilirlik (ML), maksimum sonsal (MAP (maximum a posteriori)) ve beklenen sonsal (EAP (expected a posteriori)) olmak üzere üç kestirim tekniği kullanılabilir. Ayrıca flexMIRT toplam puanları MTK ölçek puanlarına dönüştüren tablolar oluşturabilmekte ve parametre kestirimine ilişkin çeşitli standart hata kestirim tekniklerinin kullanılmasını sağlamaktadır. Bu teknikler; tamamlayıcı EM, teorik bilgi fonksiyonu, Fisher bilgi fonksiyonu ve sandwich kovaryans matrisidir. Ayrıca çıktılarında çok sayıda uyum indeksi rapor edilmektedir. Bunlardan başlıcaları, Chen ve Thissen'in χ^2 (Chen ve Thissen, 1997) Akaike Bilgi Kriteri (AIC), Bayesyen bilgi kriteridir (BIC) (Houts ve Cai, 2013).

BMIRT (Bayesian Multivariate Item Response Theory)

BMIRT programı Yao (2003) tarafından geliştirilmiştir. Programın en önemli avantajı ücretsiz olmasıdır. Programın araştırmacılar tarafından akademik çalışmalarında kullanılacağına ilişkin bir belge imzalamaları ve Java programını yüklemeleri yeterlidir. Programın kullanımına yönelik ayrıntılı bir pdf dosyası (Yao, 2013) ve program araştırmacılar tarafından Yao (2003) tarafından hazırlanmış bir web sitesinden indirilebilmektedir (<http://www.bmirt.com/6271.html>). Program kullanılmadan önce Java'nın bulunduğu yerin uzantısının bilgisayarın gelişmiş sistem ayarlarında bulunan ortam değişkenlerine yapılandırılması gerekmektedir. Program, .bat* uzantılı dosyalarla DOS komutuyla çalışmaktadır. Programı çalıştırmak için kontrol, ham veri ve uygulama olmak üzere üç farklı dosyanın hazırlanması gerekmektedir.

Bu program iki veya çok kategorili tek ve çok boyutlu modellere ilişkin parametreleri Metropolis-Hasting algoritmasını kullanan Markov zinciri Monte Carlo (Markov chain Monte Carlo (MCMC)) tekniği ile kestirmektedir. Monte Carlo yöntemleri veri simülasyonu için oldukça yaygın yöntemlerdir. MCMC kestirimleri için anahtar durum ise bilinen, uygun dağılımlardan üretilen örneklemeler kullanılarak daha karmaşık dağılımlara ilişkin örneklem üretilmesidir. Bu durum kabul-red örnekleme (Chib ve Greenberg, 1995) veya basitçe red örnekleme (Gamerman ve Lopes, 2006) olarak bilinmektedir.

BMIRT programı madde ve yetenek kestiriminin çok boyutlu ve çoklu grup veri setleri ile yapılmasına olanak tanımaktadır. Program hem açıklayıcı hem de doğrulayıcı faktör analizi, bir, iki ve üç parametrelili lojistik modeller, Rasch modeli, genelleştirilmiş iki parametrelili kısmi puan testlet modeli, aşamalı tepki modeli için parametre kestirimi yapabilmektedir (Yao, 2013).

DIMPACK V.2.0 (Parametrik olmayan Boyutluluk Değerlendirme Paketi, Nonparametric Dimensionality Assessment Package)

Program beş ayrı boyutluluk analizi yapan parametrik olmayan programı bir araya getirmiştir ve ücretsizdir. DIMPACK programı DIMTEST V.2.1, ATFINN V.1.3, DETECT V.2.1, CCPROX ve HAC programlarını bir araya getiren William Stout Enstitüsü (Nonparametric Dimensionality Assessment Package, 2006) tarafından geliştirilmiş bir paket programdır. Programın içinde bulunduğu zip dosyasını indirmek ve programla ilgili yapılmış makalelere ulaşmak için bir web sitesi bulunmaktadır (<http://psychometrictools.measuredprogress.org/dif2>). Ayrıca programın kullanımına yönelik de teknik destek verilmektedir (stoutdist@stat.uiuc.edu).

Programın kurumundan önce araştırmacıların bilgisayarlarında “NET framework” programının kurulu olduğundan emin olmaları gerekmektedir. Programla analiz yapılırken kullanılacak maksimum madde sayısı 150 ve maksimum örneklem büyüklüğü 7000’dir. Programın bir ara yüzü olduğundan kullanıcı dostudur (Deng, Han ve Hambleton, 2013).

Program boyutluluk analizinde şartlı kovaryans matrisini kullanan parametrik olmayan boyutluluk değerlendirme paketidir. DIMTEST programı verinin tek boyutlu olmadığı varsayımına dayalıdır ve maddeleri alt testlere kümelemektedir, DETECT basit yapıları için boyutluluk analizi yapmakta, CCPROX/HAC ise yakınlık matrisleri hesaplamakta ve hiyerarşik kümeleme analiziyle değerlendirme yapmaktadır. Sırasıyla bu programlar ayrı ayrı incelenmiştir.

DIMTEST

Program Stout (1987) tarafından oluşturulmuş ve daha sonra Stout, Froelich ve Gao (2001) tarafından geliştirilmiştir. DIMTEST tek boyutluluk ve yerel bağımsızlık varsayımını, parametrik olmayan hipotez ile test etmektedir (Nandakumar ve Stout, 1993; Stout, 1987; Stout ve ark., 2001). DIMTEST programının DIMPACK paketinde bulunan sürümünün yanı sıra DIMTEST_DOS şeklinde DOS sürümü de vardır. DIMTEST programında öncelikle aynı örneklemeden alınan madde havuzu AT (assessment subtest, değerlendirme alt testi) ve PT (partitioning, bölünme alt testi) olarak adlandırılan iki ayrı alt teste ayrılmaktadır. DIMPACK paketinde bulunan DIMTEST madde setlerini ikiye ayırırken DOS komutlu sürümü AT1, AT2 ve PT olmak üzere üçe ayırmaktadır. Eğer araştırmacılar DIMPACK paketi yerine DOS komutlu sürümü kullandıysa, AT1’de oluşturulan maddelerin benzer yeteneğe duyarlı oldukları varsayılmaktadır. AT2 maddelerinin ise başka bir yeteneğe, AT1 tarafından ölçülen yeteneğe göre daha duyarlı oldukları varsayımı yapılmaktadır. Ancak hem AT1 hem de AT2’deki maddelere ilişkin doğru cevap oranları dağılımlarının gözlenen frekansları açısından benzer olduğu ifade edilmektedir. Ayrıca DIMPACK paketinde bulunan DIMTEST’in avantajlarından biri de AT maddelerinin otomatik olarak seçilmesinin mümkün olmasıdır.

DIMTEST programının performansında en belirleyici olan durumlardan biri AT maddelerinin seçimidir. DIMPACK paketinde bulunan DIMTEST programında maddelerin AT ve PT şeklinde iki ayrı alt teste bölünmesi açıklayıcı veya doğrulayıcı yaklaşımla yapılabilmektedir. Doğrulayıcı yaklaşımda AT maddeleri test yapısına dayalı olan uzman görüşlerine göre oluşturulmaktadır. Açıklayıcı yaklaşımda AT maddeleri istatistiksel bir prosedürle (örneğin; madde tetrakorik korelasyonlarına ilişkin döndürülmemiş temel faktör analizi sonuçlarıyla) belirlenmektedir. DIMTEST, eğer bir test tek boyutlu ise, PT maddelerinden ayrıldıktan sonra AT alt testinde bulunan iki madde seti arasındaki şartlı kovaryansın sifıra eşit olması düşüncesine dayalıdır (Nandakumar ve Stout, 1993; Deng ve ark., 2013; Özkan ve Güvendir, 2014). AT maddelerinin belirlenmesi DETECT ve HAC/CCPROX programlarının birleşimlerini kullanan ATFINN isimli programla da mümkündür. Ayrıca alternatif olarak, araştırmacılar AT maddelerini kendileri DETECT veya HAC/CCPROX analizleri veya doğrusal olmayan faktör analizi kullanan NOHARM, TESTFACT ile veya tetrakorik korelasyon matrislerine dayalı doğrusal faktör analizi kullanarak belirleyebilirler (Nonparametric Dimensionality Assessment Package, 2006).

DIMTEST programının çıktılarında .dim* uzantılı bir dosya elde edilir. Bu dosyanın içinde AT ve PT maddeleri ile T istatistiği ve p değerleri verilmektedir. DIMTEST programıyla boyutluluk analizinde daha öncede ifade edildiği gibi bir nonparametrik hipotez vardır. Bu hipotez, “H₀: Test tek boyutludur, H₁: Test çok boyutludur” şeklinde yapılmaktadır. Bu hipotez DIMTEST tarafından hesaplanan Stout’un T istatistiği ile değerlendirilir. Eğer DIMTEST T istatistiği sıfıra yakınsa H₀ hipotezi doğrulanmış ve elde edilen p değeri manidar değilse veri kümesinin tek boyutlu olduğu sonucuna varılmaktadır. Elde edilen p değerinin manidar çıkması ise verinin çok boyutlu olduğuna işaret etmektedir (Nonparametric Dimensionality Assessment Package, 2006; Deng ve ark., 2013).

DETECT(Dimensionality Evaluation to Enumerate Contributing Traits)

DETECT uygulaması Kim (1994) ile Zhang ve Stout (1999) tarafından geliştirilmiştir. Parametrik olmayan boyutluluk analizinde şüphesiz en önemli yaklaşımlardan biridir. DETECT tekniği diğer parametrik olmayan tekniklere göre avantajlara sahiptir. Bu avantajlar sırasıyla; veri setine ait boyut sayısını kestirebilmesi, çok boyutluluk için etki büyüklüğünü hesaplayabilmesi, her bir madde tarafından baskın şekilde ölçülen boyutu belirleyebilmesidir. DETECT tekniği özellikle çok boyutlu basit test yapısını belirlemek için geliştirilmiştir (Nonparametric Dimensionality Assessment Package, 2006).

Program iki ve çok kategorili verilerle uygulanabilmekte, kayıp veriyle analiz yapamamaktadır. Eğer çok boyutlu basit yapı yaklaşık olarak ortaya çıktıysa, DETECT tekniği öncelikle tüm olası madde çiftleri arasındaki şartlı kovaryansı hesaplayarak boyutluluk analizine başlamaktadır. DETECT daha sonra testi basit yapıya ayırmayı sağlayan ortalama şartlı kovaryans matrislerini kullanarak kümeleme analizi yapmaktadır. DETECT tekniği ile boyutluluğun değerlendirilmesi için çeşitli istatistikler ve indeksler hesaplanmaktadır.

Boyutluluğun analizinde kullanılan indekslerden biri DETECT indeksidir. Bu indeksin hesaplanması için kullanılan formül aşağıda verilmiştir;

$$D(P) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \delta_{ij} E[Cov(X_i, X_j) | \Theta_{TT}]$$

Formül incelendiğinde; n madde sayısını, P , k kümeye ayrılan n madde sayısını, Θ_{TT} test birleşimini, X_i ve X_j ; i ve j maddelerine ait gözlenen puanları göstermektedir. δ_{ij} değeri eğer i ve j maddesi aynı küme içerisinde ise 1’e eşit çıkmakta değilse -1’e eşit çıkmaktadır. Teorik olarak eğer veri seti tek boyutlu ise $D(P)$ indeksinin sıfırın altında çıkması beklenmektedir. DETECT indeks değeri 1’den büyükse çok boyutluluk için güçlü bir kanıtın olduğu ifade edilmektedir. Eğer $D(P)$ 0.40 ile 1.00 arasında ise çok boyutluluk için orta düzeyde bir kanıt söz konusudur. $D(P)$, 0.20 ile 0.40 arasında ise çok boyutluluk olasılığı zayıftır ve 0.20’den küçükse verinin tek boyutlu olduğu yorumu yapılmaktadır (Zhang ve Stout, 1999; Roussos ve Ozbek, 2006; Svetina ve Levy, 2012).

DETECT tarafından hesaplanan indekslerden biri de r indeksidir. Bu indeks yaklaşık basit test yapısının yorumlanması için kullanılmaktadır. Bir çok istatistik programında testlerin basit yapıda olduğu varsayımı bulunmaktadır. Basit yapı her bir maddenin sadece tek bir boyutu ölçtüğü diğer boyutlarda hiç bir faktör yüküne sahip olmadığı çok boyutlu test yapısıdır. Ancak basit yapıya gerçek veri setlerinde çok ender rastlanmaktadır. Bu nedenle araştırmacılar tarafından yaklaşık test yapısı önerilmektedir. Yaklaşık test yapısı ise her maddenin bir boyutta baskın faktör yüküne sahip olduğu ancak diğer boyutlarda da faktör yüküne sahip olduğu çok boyutlu test yapısıdır. DETECT program ile yaklaşık test yapısı boyutluluğu incelenirken aşağıda verilen formül kullanılmaktadır:

$$r_{\max} = \frac{D(P^*)}{\hat{D}(P^*)}$$

Formül incelendiğinde, $\hat{D}(P^*)$ değeri tüm madde çiftleri boyunca şartlı kovaryans değerinden elde edilen toplam mutlak değerlerin maksimum olasılıklı değerini göstermektedir. Bu formülden elde edilen r_{max} değeri şöyle yorumlanmaktadır; eğer r indeksi 0.80 ve üstünde değerler alıyorsa verinin yaklaşık basit çok boyutlu test yapısına sahip olduğu ifade edilmektedir (Kim, 1994; Roussos ve Ozbek, 2006; Zhang ve Stout, 1999; Svetinave Levy, 2012; Jang ve Roussos, 2007).

CCPROX ve HAC

CCPROX ve HAC yaklaşımlarının ikisi birlikte açılımlı parametrik olmayan boyutluluk analizinde kullanılabilir. CCPROX/HAC testin boyutluluk yapısına ilişkin istatistiksel bilgi sağlamaktadır. CCPROX/HAC yaklaşımından testin boyutluluk yapısını belirleme sürecinde yararlanıldığı gibi DIMTEST programında kullanılan AT madde setlerinin seçimi içinde yararlanılmaktadır.

CCPROX/HAC analizi iki bölümden oluşmaktadır. HAC analizinden önce CCPROX analizinin yapılması gerekmektedir, çünkü CCPROX analiziyle oluşturulan yakınlık matrisi HAC analizinde girdi dosyası (*.prx uzantılı kaydedilmekte) olarak kullanılmaktadır. Yakınlık matrisi dosyası otomatik olarak kaydedilmekte ve eğer CCPROX analizinden hemen sonra HAC analizi yapılacaksa, dosya otomatik olarak yüklenmektedir.

HAC

HAC hiyerarşik kümeleme analizi yapmaktadır. Hiyerarşik kümeleme analizi, n sayıda her bir madde bir kümede olacak şekilde başlamakta ve daha sonra minimum yakınlığa sahip maddelerin bir araya gelmesiyle ikinci aşamaya geçilmektedir. Bu şekilde maddelerin tümünün daha büyük bir kümede yer almasıyla analiz sonuçlanmaktadır. Yeni kümeler için yakınlıklar, diğer kümelere ve daha sonra hesaplanan küme analizi türlerine (program çok sayıda alternatif sunmaktadır; karmaşık bağlantı, tek bağlantı gibi) bağlı olarak değişmektedir. Analiz hangi hiyerarşinin ya da düzeyin doğru boyut sayısı olduğu ile ilgili bilgi verememekte ancak ek tanımlayıcı istatistiksel bilgi sunmaktadır.

CCPROX

HAC analizinin amacı, test maddelerinin boyutluluk yapısı ile ilgili bilgi sunmaktır. HAC analizinin testin boyutluluğu ile ilgili kullanışlı bilgiler sunabilmesi için, madde çiftleri için hesaplanan yakınlık değerlerinin maddeler arasındaki boyutluluk farklılıklarına duyarlı olması gerekmektedir. Bu nedenle de CCPROX analizinin amacı bir yakınlık matrisi oluşturmaktır. CCPROX analizinde kullanılan yakınlık kestirimleri her bir madde çifti arasındaki şartlı kovaryans kestirimlerine dayalıdır. Eğer test çok boyutlu ise ve iki madde aynı boyutu ölçüyorsa şartlı kovaryans pozitif olmakta ve eğer maddeler farklı boyutu ölçüyorsa bu iki maddeye ilişkin şartlı kovaryans negatif olmaktadır. Bu şartlı kovaryanslardan yararlanılarak birbirine benzer boyutu ölçen maddelerin aynı kümede toplanması amaçlanmaktadır (Roussos, 1992; Roussos, Stout ve Marden, 1998).

Velicer'in Map Testi Ve Horn'un Paralel Analizi:

Paralel analiz ve Velicer'in MAP testi, özellikle SPSS ve SAS gibi yaygın kullanılan mevcut paket programlarıyla uygulanan yaklaşımlar olmaması nedeniyle araştırmacılar tarafından boyutluluk analizinde yaygın kullanılamamaktadır. Ancak özellikle paralel analiz olmak üzere bu iki tekniğin boyut sayısı belirlemede oldukça etkili yaklaşımlar oldukları yönünde literatürde görüş birliği vardır (Hayton, Allen ve Scarpello, 2004; Watkins 2006; Storch, Murphy, Bagner, Johns, Baumeister ve Goodman, 2006; Munroe ve Pearson 2006; Nelson, Canivez, Lindstrom ve Hatt, 2007; Justicia, Pichardo, Cano ve Berben, 2008; Crawford, Green, Levy, Scott, Svetina ve Thompson, 2010; Garrido, Abad ve Ponsoda, 2011).

Bu iki tekniğin uygulanabilmesi için O'Connor (2000) tarafından SPSS, SAS, MATLAB, ve R programlarıyla uygulanabilir syntax (betik) dosyaları hazırlanmıştır. Bu dosyalar ücretsiz olup, dosyalara ve kullanımına yönelik ayrıntılı bilgilerin sunulduğu bir internet sitesi mevcuttur (<https://people.ok.ubc.ca/briocconn/nfactors/nfactors.html>). Ayrıca bu betik dosyaların SPSS programında kullanımına yönelik hazırlanmış ayrıntılı videolarda bulunmaktadır (<https://www.youtube.com/watch?v=T908yGVgjPk>).

Velicer'in Map Testi

Velicer'in MAP testi, Velicer (Minimum average partial, Velicer, 1976) tarafından geliştirilmiştir. Bu teknikle boyut sayısı belirlenirken kısmi korelasyon matrislerinden ve temel bileşen analizinden yararlanılmaktadır. Boyut sayısının belirlenmek istediği maddelerden bir korelasyon matrisi oluşturulmakta ve ilk basamakta bu korelasyon matrisinden ilk boyut ayrılmaktadır. Bu işlem sırasında orijinal matristen korelasyon katsayılarının karelerinin ortalaması hesaplanmaktadır. İlk boyut ayrıldıktan sonra kalan kısmi korelasyon matrisinin köşegen elemanlarından kısmi korelasyonların karelerinin ortalamaları hesaplanmaktadır. Bu işlem değişken sayısının bir eksik basamağına kadar sürdürülebilmekte ve her basamakta yeni bir kısmi korelasyon ortalaması elde edilmektedir. Daha sonra tüm basamaklardan elde edilen kısmi korelasyonların ortalama kareleri sıralanmaktadır. Orijinal matristen elde edilen korelasyon karelerinin ortalaması; en düşük kısmi korelasyon kareler ortalamasından daha küçük olduğu anda, artık matristen herhangi bir bileşenin/faktörün ayıramayacağı ve boyut sayısının o değere kadar olan basamak sayısı kadar olması gerektiği önerilmektedir. Diğer bir deyişle analizde en düşük kısmi korelasyon ortalama karesinin basamak numarası bileşen sayısı olarak ifade edilmektedir (Velicer, 1976; Velicer, Eaton ve Fava, 2000; O'Connor, 2000; Watkins, 2006; Garrido ve ark., 2011).

Paralel Analiz

Paralel analiz, Horn (1965) tarafından boyut sayısını belirlemeye yönelik geliştirilen bir tekniktir. Paralel analizde gerçek veriyle aynı özelliklere sahip tesadüfi bir seri korelasyon matrisi oluşturulmakta ve hem bu rastgele oluşturulmuş korelasyon matrislerinden hem de gerçek veriden temel bileşen analiziyle özdeğerler hesaplanmaktadır. Boyut sayısı belirlenirken gerçek veriden hesaplanan öz değerlerin tesadüfi verilerden hesaplanan öz değerden büyük olduğu nokta referans olarak alınmaktadır. Diğer bir deyişle gerçek veriden hesaplanan i. sıradaki öz değer, tesadüfi veriden hesaplanan i. sıradaki öz değerden daha büyük olduğu basamak sayısı boyut sayısını göstermektedir (Zwick ve Velicer, 1986; O'Connor, 2000; Watkins, 2006; Ladesma ve Valero-Mora, 2007; Piconne, 2009).

Programlara Yönelik Yapılmış Araştırmalar

Knol ve Berger (1991) tarafından TESTFACT, NOHARM gibi çok boyutlu MTK programları ile beş farklı faktörleme tekniğini bir simülasyon çalışmasında karşılaştırmış ve farklı test koşullarında TESTFACT ile NOHARM programlarının benzer performansa sahip olduğunu ancak açık şekilde MAXLOG'a üstün olduğunu ifade etmişlerdir.

Gosz ve Walker (2001), basit ve karmaşık test yapılarıyla TESTFACT ve NOHARM programlarını simülasyon çalışmasıyla karşılaştırmışlardır. NOHARM programının TESTFACT'e göre madde performansı açısından daha iyi sonuçlar verdiğini ifade etmişlerdir.

Tate (2003), Mplus, TESTFACT ve NOHARM programlarıyla boyutluluk analizi yapmış; hem gerçek verilerle hem de simülasyon verileriyle bu programları karşılaştırmıştır. Hem gerçek veri setleriyle hem de üretilmiş veri setleriyle üç programın da benzer faktör yapılarını ortaya çıkardığı ifade edilmiştir.

Stone ve Yeh (2006), çoktan seçmeli test maddelerinin boyutluluğunun değerlendirilmesinde Mplus, NOHARM ve TESTFACT programlarını karşılaştırmıştır. Çalışmadan elde edilen sonuçlara göre şansla cevaplama olasılığına sahip veri setlerine yönelik üç programda benzer faktör çözümleri üretmiştir.

Finch ve Habing (2005), DETECT ve NOHARM programlarını bir simülasyon çalışmasıyla karşılaştırmışlardır. Farklı test koşulları (örtük yetenek dağılımları, test uzunluğu, örneklem büyüklüğü, boyutlar arasındaki korelasyon) kullanılarak bir simülasyon çalışmasıyla DETECT ve NOHARM programının maddeleri kümeleme, ve hangi maddenin hangi boyutta yer aldığını belirleme özellikleri araştırmacılar tarafından incelenmiştir. Araştırmacılar, NOHARM programının boyut sayısını daha doğru belirlediğini, ancak hangi maddenin hangi boyutta yer aldığını DETECT programına göre daha hatalı belirlediğini ifade etmişlerdir.

Finch ve Habing (2007) tarafından DIMTEST ve NOHARM programlarının sonuçları, farklı test uzunluğu, farklı örneklem büyüklükleri, şansla cevaplama olasılığının varlığı ve yokluğu, farklı yetenek dağılımları değişkenleri kapsamında incelenmiştir. DIMTEST ile NOHARM'a dayalı istatistiklerin birinci tip hatalarının düşük olduğu, benzer ki-kare istatistiği değerlerine sahip oldukları görülmüştür. Ancak şansla cevaplama olasılığının söz konusu olduğu durumlarda DIMTEST programının daha düşük birinci tip hataya sahip olduğu, diğer bir deyişle daha iyi sonuçlar ürettiği bulgusuna ulaşılmıştır.

Gessaroli ve Champlain (1996) tarafından, NOHARM ve DIMTEST programları farklı örneklem büyüklüğü (500, 1000) ve test uzunluğu (15, 30, 45) gibi test koşullarıyla karşılaştırılmıştır. Çalışmalarında test koşullarının çoğunda NOHARM programının ürettiği yaklaşık ki-kare istatistiğinin çoğu durumda DIMTEST programının ürettiği Stout'un T istatistiğine benzer sonuçlar verdiği, ancak küçük örneklem büyüklüğü ve test uzunluğuyla NOHARM programının daha iyi sonuçlar ürettiği ifade edilmiştir.

Yeh (2007) tarafından şans başarısının mümkün olduğu çoktan seçmeli test maddeleriyle Mplus ve TESTFACT programları kullanılarak madde ayırt edicilik düzeyleri ve boyutlar arasındaki korelasyonda manipüle edilerek boyutluluğun analizinde kullanılan dört indeks (varyans oranı, paralel analiz, RMSR indirgeme katsayısı ve ki-kare farklılık testi) karşılaştırılmıştır. Çalışmada hem simülasyon hem de TIMSS 2003 verisiyle uygulama yapılmıştır. Araştırmada, TESTFACT programının şansla cevaplamanın söz konusu olduğu çoğu koşulda Mplus programına göre daha iyi sonuç verdiği, RMSR indirgeme indeksinin ve varyans oranının Mplus ile daha doğru kestirildiği ancak ki-kare testi ve paralel analizin TESTFACT ile daha doğru uygulandığı bulgusuna ulaşılmıştır. Çalışmada boyutluluk analizinde TESTFACT programının Mplus programına göre daha tutarlı sonuçlar ürettiği ifade edilmiştir.

Asparouhov ve Muthen (2012) tarafından, Mplus ve IRTPRO programları kullanılarak açımlayıcı ve doğrulayıcı faktör analizi için kestirim teknikleri (maksimum olabilirlik, MH-RM, Bayes tekniği) simülasyon çalışmasıyla karşılaştırılmıştır. Ortalama yanlılık ve hata değerlerinin iki program ve programlara ilişkin kestirim teknikleri açısından hem iki faktörlü doğrulayıcı model için hem de açımlayıcı faktör analiziyle birbirine oldukça yakın değerler aldığı bulgusuna ulaşılmıştır.

Yavuz (2014) tarafından, BMIRT programıyla madde parametresi kestiriminde kullanılan MCMC tekniği, flexMIRT program tarafından kullanılan BA-EM ve MH-RM teknikleri çeşitli test koşullarıyla (örneklem büyüklüğü, test uzunluğu, boyutlar arasındaki korelasyon ve boyut sayısı) karşılaştırılmıştır. Özellikle büyük örneklerle ve test uzunluklarıyla iki programında çok düşük hatalarla madde parametresi kestirimi yaptığı, boyut sayısı fazla olduğunda (örneğin 5 iken) flexMIRT ile uygulanan BA-EM tekniğinin madde parametresi kestirimi yapmadığı, MH-RM ve MCMC tekniğinin benzer doğrulukta kestirimler yaptığı bulgusuna ulaşılmıştır. Özellikle örneklem büyüklüğünün çok boyutlu verilerle madde parametresi kestiriminde önemli olduğu bulgusuna ulaşılmıştır.

Hattie, Krakowski, Rogers ve Swaminathan (1996) tarafından, DIMTEST programıyla hesaplanan Stout'un T istatistiği tek boyutlu ve çok boyutlu verilerle farklı test koşullarında incelenmiş ve bu istatistiğin telafi edici ve yaklaşık telafi edici çok boyutlu modellerde duyarlı ölçüm yaptığı bulgusuna ulaşılmıştır. Mearai Robin ve Sireci (2000) ise testteki madde sayısı az olduğunda boyutluluğun analizinde DIMTEST'in hatalı sonuçlar ürettiğini ifade etmiştir. DIMTEST programının 25'den küçük maddeye sahip testlerle ve 500'den küçük örneklem büyüklükleriyle kötü

sonuçlar verdiği farklı araştırmacılar tarafından da ifade edilmiştir (De Champlin ve Gessaroli, 1991; De Champlin ve Tang, 1993; Finch ve Habing, 2007).

Svetina (2011) tarafından, telafi edici ve telafi edici olmayan çok boyutlu madde tepki modellerini karmaşık test yapısı ile incelenmiş ve bu modellerin boyutluluğunu değerlendirmek için şartlı kovaryanslara dayalı DETECT ve faktör analitik yaklaşımlara dayalı NOHARM programlarını karşılaştırmıştır. Bu araştırmanın sonuçlarına göre iki ve üç boyutlu telafi edici modeller için DETECT temelli tekniklerin NOHARM'a göre daha iyi sonuç verdikleri ifade edilmiştir. Araştırmacı tarafından DETECT temelli tekniklerin özellikle boyutlar arasındaki korelasyon 0.60 ve altında olduğu durumlarda, model %30 ve daha düşük oranda karmaşık olduğunda daha iyi sonuç ürettiği bulgusuna ulaşmıştır. Ayrıca DETECT temelli tekniklerin örneklem büyüklüğü arttıkça daha iyi sonuçlar ürettiği ifade edilmiştir. Ancak modele ilişkin test yapısı karmaşık hale geldikçe DETECT programı ile kötü sonuçlar elde edildiği bulgusuna ulaşılmıştır. DETECT programının özellikle basit test yapısındaki maddelerin sınıflandırılmasında tutarlı sonuçlar ürettiği, karmaşık test yapısındaki maddelerin sınıflandırılmasında daha az tutarlı sonuçlar ürettiği ifade edilmiştir.

Meara, Robin ve Sireci (2000) tarafından, çok boyutlu ölçekleme yaklaşımı ile DIMTEST programı iki kategorili verilerin boyutluluğunu değerlendirmek üzere karşılaştırılmıştır. Test uzunlukları ve boyutlar arasındaki korelasyonlar farklılaştırılarak tek ve çok boyutlu veri setleri üretilmiş ve çalışmada elde edilen en önemli bulgulardan biri DIMTEST programının çoğu test koşulunda iyi sonuçlar verdiği ancak testteki madde sayısı az olduğunda hatalı sonuç ürettiğidir.

Deng ve Ansley (2000) tarafından telafi edici ve telafi edici olmayan çok boyutlu modellerin belirlenmesi için DIMTEST programı kullanılmıştır. Örneklem büyüklüğü arttıkça ve boyutlar arasındaki korelasyon azaldıkça telafi edici olmayan çok boyutluluğu daha doğru belirlediği ifade edilmiştir. Araştırmacılar DIMTEST'in telafi edici çok boyutlu veri setlerinin boyutluluğunu belirlemede hatalı sonuçlar ürettiğini ifade etmiştir.

SONUÇLAR ve TARTIŞMA

Bu çalışmada çok boyutlu veri setleri ve modelleri ile boyutluluk analizini yapma olanağı tanıyan çeşitli paket programlar tanıtılmıştır. Programların ulaşılabilirlikleri, ne tür veri setleri ve modellerle analiz yapabildikleri, programların teknik özellikleri, çıktı dosyalarında bulunan indekslerin nasıl yorumlandıkları gibi programlara ait önemli özellikler açıklanmaya çalışılmıştır. Son olarak programların performanslarını karşılaştıran çeşitli araştırma sonuçlarına yer verilmiştir. İnceleme sonuçlarına göre programların çok sayıda özellik açısından farklılaştıkları, avantajlı ve dezavantajlı yönlerinin olduğu söylenebilir. Örneğin araştırmacıların ve uygulayıcıların veri setlerinde kayıp veri olması durumunda bazı programlarla analizler yapma olanağı varken diğerlerinde yoktur. Araştırmacıların programların maliyetleri konusunda da fikir sahibi olmaları önemlidir ve bazı programları (NOHARM, DIMPACK (DETECT, DIMTEST, CCPROX/HAC), BMIRT, Paralel Analiz, Velicer'in MAP Testi) ücretsiz edinmek mümkünken bazıları (flexMIRT, TESTFACT, Mplus, IRTPRO) ücretlidir. Ayrıca veri setlerin iki ve çok kategorili olup olmaması, örneklem büyüklüğü önemli değişkenlerdir. Eğer veri setleri parametrik olmayan özellikte iseler tercih edilmesi gereken programlar da parametrik olmayan yaklaşımlara dayalı olan programlardır (DETECT, DIMTEST, CCPROX/HAC). Program seçiminde önemli ölçütlerden biri de programların kullandıkları kestirim teknikleridir. Örneğin, eğer Bayesyen yaklaşıma dayalı bir kestirim tekniği tercih edildiyse (örneğin BMIRT programı ile uygulanabilen MCMC) bu tekniğin kestirim için daha uzun süreler gerektirdiği göz önünde bulundurulmalıdır. Öte taraftan programların kullanım koşulları da önemlidir, örneğin dos komutuyla çalışan veya yazılım dili zor olan ve ara yüzü olmayan bir programa aşına olmayan araştırmacılar programı kullanma konusunda zorluk yaşayabilirler, bu durumda ara yüzü olan bir programların kullanılması kolaylık sağlayacaktır. Son olarak hangi test koşullarında hangi programın performansı daha iyidir sorusunun cevabı çeşitli simülasyon çalışmalarında bulunmaktadır. Bu nedenle araştırmacıların test koşullarına yönelik yapılmış simülasyon çalışmalarına yönelik iyi bir alanyazın çalışması yapmaları katkı sağlayıcı olabilmektedir.

KAYNAKÇA

- Ackerman, T. A. (1989). Unidimensional IRT calibration of compensatory and noncompensatory multidimensional items. *Applied Psychological Measurement, 13*, 113-127.
- Ackerman, T. A. (1996) Graphical representation of multidimensional item response theory analyses. *Applied Psychological Measurement, 20*, 311-329.
- Ackerman, T. A., Gierl, M. J., & Walker, C.M. (2003) Using multidimensional item response theory to evaluate educational and psychological tests. *Educational Measurement: Issues and Practice, 22*, 37-51.
- Adams, R. J., Wilson, M. R., & Wang, W. C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement, 21*, 1-23.
- Ansley, R. A., & Forsyth, T. N. (1985). An examination of the characteristics of unidimensional IRT parameter estimates derived from two-dimensional data. *Applied Psychological Measurement, 9*, 37-48.
- Asparouhov, T. & Muthén, B. (2012). Using Mplus TECH11 and TECH14 to test the number of latent classes. *Mplus Web Notes, 14*, 22.
- Beguín, A. A., & Glas, C. A. W. (2001). MCMC estimation and some model fit analysis of multidimensional IRT models. *Psychometrika, 66*, 541-562.
- Bentler, P.M., & Bonnet, D.C. (1980). Significance Tests and Goodness of Fit in the Analysis of Covariance Structures. *Psychological Bulletin, 88* (3), 588-606.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Applications of an EM algorithm. *Psychometrika, 46*, 443-459.
- Bock, R. D., Gibbons, R., & Muraki, E. (1988) Full information item factor analysis. *Applied Psychological Measurement, 12*, 261-280.
- Bock, R. D., & Schilling, S. G. (2003). IRT based item factor analysis. In M du Toit (ed) *IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFACT*, 584-591. Scientific Software International, Lincolnwood, IL.
- 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.
- Bolt, D. M., & Lall, V. F. (2003). Estimation of compensatory and noncompensatory multidimensional item response models using Markov chain Monte Carlo. *Applied Psychological Measurement, 29*, 395-414.
- Bradlow, E. T., Wainer, H., & Wang, X. (1999). A Bayesian random effects model for testlets. *Psychometrika, 64*, 153-168.
- Byrne, B.M. (1998). *Structural Equation Modeling with LISREL, PRELIS and SIMPLIS: Basic Concepts, Applications and Programming*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Cai, L., du Toit, S. H. C., & Thissen, D. (2011). *IRTPRO: Flexible, Multidimensional, Multiple Categorical IRT Modeling*. Scientific Software International.
- Cai, L. (2013). *flexMIRT version 2.00: A numerical engine for flexible multilevel multidimensional item analysis and test scoring. [Computer software]*. Chapel Hill, NC: Vector Psychometric Group.
- Chen, W., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics, 22*, 265-289.
- Chib, S. & Greenberg, E. (1995). Understanding the Metropolis Hastings Algorithm. *American Statistical Journal, 49*, 327-335.
- Crawford, A. V., Green, S. B., Levy, R., Lo, W. J., Scott, L., Svetina, D., & Thompson, M. S. (2010). Evaluation of parallel analysis methods for determining the number of factors. *Educational and Psychological Methods, 70*, 885-901.
- DeChamplain, A. F. & Gessaroli, M. E. (1991, April). *Assessing test dimensionality using an index based on non-linear factor analysis*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL. (ERIC Document Reproduction Service No. 334 235).
- DeChamplain, A. & Tang, L. (1993, April). *The effect of non-normal ability distributions on the assessment of dimensionality*. Paper presented at the annual meeting of the National Council on Measurement in Education, Atlanta, GA.
- De Champlain, A. F. (1996). The effect of multidimensionality on IRT true-score equating for subgroups of examinees. *Journal of Educational Measurement, 33*, 181-201.
- Deng, H. & Ansley, T. (2000, April). *Detecting compensatory and noncompensatory multidimensionality using DIMTEST*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA (ERIC Document Reproduction Service No. 445 029. Retrieved from <http://catalogue.nla.gov.au/Record/5673279>).

- Deng, N., Han, K., T. & Hambleton, R., K., (2013). A Review of DIMPACK Version 1.0: Conditional Covariance Based Test Dimensionality Analysis Package. *Applied Psychological Measurement*, 37 (2), 162-172.
- Diamantopoulos, A. and Siguaw, J.A. (2000). *Introducing LISREL*. London: Sage Publications.
- Drasgow, F., & Parsons, C. K. (1983). Application of unidimensional item response theory models to multidimensional data. *Applied Psychological Measurement*, 7, 189-199.
- du Toit, M. (2003). *IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFACT*. Lincolnwood, IL: Scientific Software International.
- Embretson, S. E. (1997). Multicomponent response models. In W. J. Van der Linden R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 305–321). New York: Springer Verlag.
- Finch, H., & Habing, B. (2005). Comparison of NOHARM and DETECT in item cluster recovery: Counting dimensions and allocating items. *Journal of Educational Measurement*, 42, 149-169.
- Finch, H., & Habing, B. (2007). Performance of DIMTEST- and NOHARM-based statistics for testing unidimensionality. *Applied Psychological Measurement*, 31, 292-307.
- Fraser, C. (1988). *NOHARM II: A Fortran program for fitting unidimensional and multidimensional normal ogive models of latent trait theory* [Software]. Armidale, New South Wales: University of New England, Centre for Behavioral Studies.
- Fraser, C., & McDonald, R. P. (1988). NOHARM: Least squares item factor analysis. *Multivariate Behavioral Research*, 23, 267-269.
- Gamerman, D. & Lopes, H. F. (2006) *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Second Edition. London: Chapman & Hall/CRC Press.
- Garrido, L. E., Abad, F. J., & Ponsoda, V. (2011). Performance of Velicer's Minimum Average Partial Factor Retention Method with Categorical Variables. *Educational and Psychological Measurement*, 71 (3), 551-570.
- Gelman, A., Meng, X. L. & Stern, H. S. (1996). Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statist. Sinica* 6, 733–760.
- Gessaroli, M. E., & De Champlain, A. F. (1996). Using an approximate chi-square statistic to test the number of dimensions underlying the responses to a set of items. *Journal of Educational Measurement*, 33, 157-192.
- Gosz, J. & Walker, C. M. (2001). *An Empirical Comparison of Multidimensional Item Response Data Using TESTFACT and NOHARM*. Annual meeting of the American Educational Research Association, New Orleans, LA, USA.
- Hattie, J., Krakowski, K., Rogers, J., & Swaminathan, H. (1996). An assessment of Stout's index of essential dimensionality. *Applied Psychological Measurement*, 20, 1-14.
- Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, 7 (2), 191-205.
- Harrison, D. A. (1986). Robustness of IRT parameter estimation to violations on the unidimensionality assumption. *Journal of Educational Statistics*, 11, 91–115.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30 (2), 179-185.
- Houts, C. R., & Cai, L. (2013). *flexMIRT user's manual version 2.0: flexible multilevel multidimensional item analysis and test scoring*. Chapel Hill, NC: Vector Psychometric Group.
- Hu, L.T., & Bentler, P.M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives. *Structural Equation Modeling*, 6 (1), 1-55.
- Jang, E. E., & Roussos, L. A. (2007). An investigation into the dimensionality of TOEFL using conditional covariance-based non-parametric approach. *Journal of Educational Measurement*, 44 (1), 1-21.
- Jöreskog, K., & Sörbom, D. (1993). *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*. Chicago, IL: Scientific Software International Inc.
- Justicia, F., Pichardo, M. C., Cano, F., Berben, A. B. G., & De la Fuente, J. (2008). The revised two-factor study process questionnaire (R-SPQ-2F). Exploratory and confirmatory factor analyses at item level. *European Journal of Psychological of Education*, 23 (3), 355-372.
- Kim, H.R. (1994). *New techniques for the dimensionality assessment of standardized test data*. (Doctoral Dissertation, University of Illinois at Urbana—Champaign).
- Kenny, D.A., & McCoach, D.B. (2003). Effect of the Number of Variables on Measures of Fit in Structural Equation Modeling. *Structural Equation Modeling*, 10 (3), 333-51.
- Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling* (2nd ed.). New York: Guilford. 366 pp., ISBN 978-1-57230-690-5.
- Knol, D. L. & Berger, M. P. (1991). Empirical comparison between factor analysis and multidimensional item response models. *Multivariate Behavioral Research*. 26 (3), 457-477.

- Ladesma, R.D., & Valero- Mora, P. (2007). Determining the number of factors to retain in EFA: An easy-to-use computer program for carrying out parallel analysis. *Practical Assessment, Research and Evaluation, 12*, 1-11.
- MacCallum, R.C., Browne, M.W., & Sugawara, H., M. (1996). Power Analysis and Determination of Sample Size for Covariance Structure Modeling. *Psychological Methods, 1* (2), 130-49.
- McDonald, R. P. (1962). A general approach to nonlinear factor analysis. *Psychometrika, 27*, 398-415.
- McDonald, R. P. (1967). *Nonlinear factor analysis [Psychometric Monographs, No. 15]*. Chicago: University of Chicago Press.
- McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology, 34*, 100-117.
- McDonald, R. P. (1997). *Normal-ogive multidimensional model*. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 257-269). New York, NY: Springer-Verlag.
- McDonald, R. P. (2000). A basis for multidimensional item response theory. *Applied Psychological Measurement, 24*, 99-114.
- Meara, K., Robin, F., & Sireci, S.G. (2000). Using Multidimensional Scaling to Assess the Dimensionality of Dichotomous Item Data, *Multivariate Behavioral Research, 35* (2), 229–259.
- Munroe, A., & Pearson, C. (2006). The Munroe Multicultural Attitude Scale Questionnaire: A new instrument for multicultural studies. *Educational and Psychological Measurement, 66*, 819-834.
- Muthén, L.K. & Muthén, B.O. (1998-2012). *Mplus User's Guide. Seventh Edition*. Los Angeles, CA: Muthén & Muthén.
- Nelson, J. M., Canivez, G. L., Lindstrom, W., & Hatt, C. V. (2007). Higher-order exploratory factor analysis of the Reynolds Intellectual Assessment Scales with a referred sample. *Journal of School Psychology, 45*, 439–456.
- Nandakumar, R. & Stout, W. (1993). Refinements of stout's procedure for assessing latent trait unidimensionality. *Journal of Educational Statistics, 18* (1), 41-68.
- Nonparametric Dimensionality Assessment Package (2006)., Champaign, IL: William Stout Institute for Measurement.
- O'Connor, B.P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers, 32*, 396-402.
- Özer Özkan, Y., & Acar Güvendir M. (2014). Türkiye'de Uygulanan Geniş Ölçekli Testlerin Çok Boyutluluğunun Analizi. *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi, 29*, 31-47.
- Piccone, A.V. (2009). *A comparison Of Three Computational Procedures for Solving the Number of factors problem in exploratory factor analysis*. (Doctoral dissertation. University of Northern Colorado).
- Reckase, M. D. (1997). The past and future of multidimensional item response theory. *Applied Psychological Measurement, 27*, 25-36.
- Reckase, M. D. (2009). *Multidimensional Item Response Theory*. Springer-Verlag, New York.
- Roussos, L. A. (1992). *Hierarchical agglomerative clustering computer program user's manual*. (Doctoral Dissertation, University of Illinois, Urbana-Champaign).
- Roussos, L. A., Stout, W. F., & Marden, J. I. (1998). Using new proximity measures with hierarchical cluster analysis to detect multidimensionality. *Journal of Educational Measurement, 35*, 1-30.
- Roussos, L.A. & Ozbek, O. (2006). Formulation of the DETECT population parameter and evaluation of DETECT estimator bias. *Journal of Educational Measurement, 43*, 215-243.
- Sireci, S. G., Thissen, D., & Wainer, H. (1991). On the Reliability of Testlet-Based Tests. *Journal of Educational Measurement, 28* (3), 237-247.
- Shealy, R., & Stout, W. (1993). A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DIF as well as item bias/DIF. *Psychometrika, 58*, 159-194.
- Steiger, J.H. (1990). Structural model evaluation and modification. *Multivariate Behavioral Research, 25*, 214-12.
- Steiger, J.H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences, 42* (5), 893-98.
- Stone, C. A., & Yeh, C. C. (2006). Assessing the dimensionality and factor structure of multiple-choice exams: An empirical comparison of methods using the Multistate Bar Examination. *Educational and Psychological Measurement, 66*, 193–214.
- Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika, 52*, 589-617.
- Stout, W., Froelich, A. G., & Gao, F. (2001). *Using resampling methods to produce an improved DIMTEST procedure*. In A. Boomsma, M. A. J. van Duijn, & T. A. B. Snijders (Eds.), *Essays on item response theory* (pp. 357-375). New York, NY: Springer-Verlag.

- Stout, W. (2006). *DIMPACK 1.0*. Chicago, IL: Assessment Systems Corporation.
- Storch, E. A., Murphy, T. K., Bagner, D. M., Johns, N., Baumeister, A., & Goodman, W. K. (2006). Reliability and Validity of the child behavior checklist obsessive-compulsive scale. *Psychiatry Research, 129*, 91-98.
- Svetina, D. (2011). *Assessing Dimensionality in Complex Data Structures: A Performance Comparison of DETECT and NOHARM Procedures*. (Doctoral dissertation, Arizona State University). Retrieved from <http://repository.asu.edu/attachments/56763/content>.
- Svetina, D., & Levy, R. (2012). An Overview of Software for Conducting Dimensionality Assessment in Multidimensional Models. *Applied Psychological Measurement, 36* (8), 659-669.
- Tabachnick, B.G., & Fidell, L.S. (2007). *Using Multivariate Statistics (5th ed.)*. New York: Allyn and Bacon.
- Tate, R. (2003). A comparison of selected empirical methods for assessing the structure of responses to test items. *Applied Psychological Measurement, 27*, 159-203.
- Velicer, W.F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 41*, 321-327.
- Velicer, W. F., Eaton, C.A., & Fava, J.L. (2000). *Construct Explication through Factor or Component Analysis: A Review and Evaluation of Alternative Procedures for Determining the Number of Factors or Components*. In Goffin, R. D., & Helmes, E. (Eds.), *Problems and Solutions in Human Assessment: Honoring Douglas Jackson at Seventy*. Boston: Kluwer. (pp. 41-71).
- Walker, C. M., & Beretvas, S. N. (2003). Comparing multidimensional and unidimensional proficiency classifications: Multidimensional IRT as a diagnostic aid. *Journal of Educational Measurement, 40*, 255-275.
- Watkins, M. W. (2006). Determining parallel analysis criteria. *Journal of Modern Applied Statistical Methods, 5*, 344-346.
- Way, W. D., Ansley, T. N., & Forsyth, R. A. (1986). *The effects of two-dimensional data on unidimensional IRT parameter estimates*. Annual meeting of the American Educational Research Association, San Francisco CA.
- Wheaton, B., Muthen, B., Alwin, D., F., & Summers, G. (1977). Assessing Reliability and Stability in Panel Models. *Sociological Methodology, 8* (1), 84-136.
- Yavuz, G. (2014). Çok Boyutlu Madde Tepki Kuramı Modelleri ve Programları için Karşılaştırmalı Analizler. (Yayımlanmamış Doktora tezi, Hacettepe Üniversitesi, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Ankara).
- Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement, 30*, 187-213.
- Yao, L. (2003). *BMIRT: Bayesian multivariate item response theory*. CTB/McGraw-Hill, Monterey, CA.
- Yao, L., & Boughton, K. A. (2009). Multidimensional Linking for Tests with Mixed Item Types. *Journal of Educational Measurement, 46* (2), 177-197.
- Yao, L. (2013). *The BMIRT Toolkit*. Defense Manpower Data Center, DoD Center Monterey Bay, US.
- Yeh, C. (2007). *The effect of guessing on assessing dimensionality in multiple-choice tests: A Monte Carlo study with application*. (Doctoral Dissertation. University of Pittsburg). Retrieved from <http://d-scholarship.pitt.edu/7286/>.
- Zhang, J., & Stout, W. (1999). The theoretical DETECT index of dimensionality and its application to approximate simple structure. *Psychometrika, 64*, 213-249.
- Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99*, 432-442.

EXTENDED ABSTRACT

The assumption of unidimensionality tests and test items has become more questionable after 1980s (Ackerman, 1989; Ansley & Forsyth, 1985; Drasgow & Parsons, 1983; Harrison, 1986; Way, Ansley & Forsyth, 1986). Especially in recent years, there is substantial demand for multidimensional data sets and their various test applications (such as equating test scores, subtest score reporting, differential item functioning). There are various software currently available for assessment of dimensionality. The researchers, who want to make their analysis with multidimensional data sets and models, must to review the software packages in terms of their capabilities. There are important differences between software packages such as their technical details, their capabilities to conduct dimensionality

assessment, types of data, types of models, estimation techniques, estimation time, availability of them.

In this study, some popular software packages (Mplus, NOHARM, TESTFACT, IRTPRO, flexMIRT, BMIRT, DIMTEST, DETECT, CCPROX/HAC, Velicer's MAP test and Parallel Analysis) which are used for analyzing test dimensionality and used for various test applications with multidimensional data sets, was investigated. The software packages was investigated in terms of their availability, their capabilities of data sets, models, estimation techniques. The software packages' dimensionality analyses approach (confirmatory, exploratory) was investigated. Also assessing of the indices which can be gained from software packages' outputs was investigated. Finally the simulation studies in which the packages are compared in terms of their various properties was reviewed.

Mplus software was developed by Muthen (1998-2012). There are a pdf file and website which include using and availability of the software (<https://www.statmodel.com/ugexcerpts.shtml>). Mplus offers researchers a wide choice of models, estimators, and algorithms in a program that has an easy-to-use interface and graphical displays of data and analysis results. Mplus allows the analysis of both cross-sectional and longitudinal data, single-level and multilevel data and data that come from different populations with either observed or unobserved heterogeneity Mplus has extensive capabilities for Monte Carlo simulation studies, where data can be generated and analyzed according to any of the models included in the program. Mplus can be used in either exploratory and confirmatory factor analysis.

NOHARM (Normal Ogive Harmonic Analysis Robust Method) was developed by McDonald (1967) as a model and was developed as software package by Fraser (1998). The detailed information about using software can be accessed with the papers which has written by McDonald (1962; 1967; 1981; 1997; 2000). NOHARM can be used with unidimensional and multidimensional normal ogive models. NOHARM contains options for orthogonal and oblique rotations for exploratory factor solutions and it can be used in confirmatory factor analysis. NOHARM output for confirmatory factor analysis of dimensionality includes RMSR and Tanaka's goodness-of-fit index.

TESTFACT (Full Information Item Factor Analysis) was developed by Bock, Gibbons and Muraki (1998) as a model and the 4.0 version of TESTFACT package was developed by Gibbons, Schilling, Muraki, Wilson and Wood (2003). TESTFACT supports analyses of polytomous and dichotomous data, missingness and proceeds by computing tetracoric correlation matrix. TESTFACT contains options for orthogonal and oblique rotations for exploratory factor solutions and it can be used in confirmatory factor analysis.

IRTPRO was developed by Cai, du Toit and Thissen (2011). IRTPRO offers exploratory and confirmatory analysis with a large number of unidimensional and multidimensional models (One, two and three parameter logistic models, Graded model, Generalized Partial Credit and Nominal) and data sets.

flexMIRT 2.0 was developed by Cai (2013). flexMIRT has some of the richest psychometric and statistical features but it is not free available. flexMIRT can estimate item parameter for a wide range of unidimensional and multidimensional models (Rasch, one, two, three parameter logistic models, Graded model, Generalized Partial Credit and Nominal) with exploratory and confirmatory approach. flexMIRT fits a variety of unidimensional and multidimensional IRT models to single-level and multilevel data using maximum marginal likelihood or modal Bayes via Bock-Aitkin EM (with generalized dimension reduction) or MH-RM estimation algorithms.

BMIRT was developed by Yao (2003). One of the advantage of the software is that it is free available. BMIRT (Yao, 2003) (Bayesian Multivariate Item Response Theory) that implements Markov Chain Monte Carlo (MCMC) methods using the Metropolis–Hastings sampling algorithm to estimate item, examinee, and population distribution parameters for a set of MIRT models for both dichotomously and polytomously scored test items. Both confirmatory and exploratory item factor analysis are possible. The program can perform unidimensional or multidimensional calibrations. It can operate on a single group or multiple groups. It can fit dichotomous or polytomous models (along with mixed models), including the three-parameter logistic model, the two-parameter logistic model, the Rasch model, the generalized two-parameter partial credit model, the testlet model, the graded response model, and the higher-order IRT model.

Nonparametric Dimensionality Assessment Package contains 5 different programs (DIMTEST v.2.1, ATFIND v.1.3, DETECT v.2.1, CCPROX and HAC) and it was developed by William Stout Institute (2006). DIMTEST is a nonparametric hypothesis testing procedure to test the assumptions of unidimensionality and local independence. When simple structure multidimensionality is even just approximately present, the DETECT procedure allows for a complete nonparametric description of the multidimensionality. The DETECT procedure begins by calculating the mean conditional covariance between all possible pairs of test items. CCPROX and HAC can be used together to conduct an exploratory nonparametric dimensionality analysis. CCPROX/HAC analysis computes proximity matrices and assesses dimensionality with hierarchical cluster analysis.

Although they are highly recommended by researchers and recent studies, popular software packages (such as SPSS, SAS), can not permit analysis of dimensionality with parallel analysis and Velicer's MAP (manimum average partial) test directly. Because of this the using of them are not very common with dimensionality analysis. The MAP test and parallel analysis can be implemented with the use of available syntax (O'Connor, 2000).

In summary, there are important differences between software package which can be used for assessing dimensionality, it appeared that each of the software packages had both advantages and disadvantages, and that users need to show care in making their choices. The researchers must know the software capabilities (such as estimation technique, type of data, models, availability, estimation technique and estimation time).

Ekler

Ek 1. Programların Karşılaştırılması

Bilgisayar Programları	Erişim	Web sitesi	Veri Türleri	Kestirim Teknikleri	Boyutluluk Analizi	Sınırlılığı
Mplus	Ücretli	https://www.statmodel.com/ug excerpts.shtml	-Tek ve çok düzeyli veri, kayıp veri, her türlü ölçekten elde edilmiş veri setleri ve kombinasyonları -Monte Carlo Simülasyonu	-Bayesyen ve olabilirlik kestirimleri	-Açımlayıcı, Doğrulayıcı	-Mplus için komut yazılması gerekmekte
NOHARM	Ücretsiz	http://noharm.software.informer.com/download/	-İki Kategorili	-Polinomial ve ağırlıklandırılmamış en küçük Kareler Yöntemi	-Açımlayıcı, Doğrulayıcı	-Şans ve yetenek parametresi kestirimi yapamamaktadır, Çok kategorili ve kayıp verilerle çalışmamaktadır
TESTFACT	Ücretli	http://www.ssicentral.com/irt/index.html	- İki ve çok kategorili veriler, kayıp veriye sahip veri setleri	- Marjinal maksimum olabilirlik kestirimini ve ayrıca yetenek parametreleri için Bayesyen algoritması	-Açımlayıcı, Doğrulayıcı (sadece iki faktörlü)	-Şans parametresi kestirimi yapamamaktadır
IRTPRO	Ücretli	http://www.ssicentral.com/irt/index.html	- İki ve çok kategorili veriler, kayıp veriye sahip veri setleri	-Maksimum olabilirlik, MAP, Bock Aitkin EM Bifactor EM, Generalized Dimension Reduction EM, MHRM	-Açımlayıcı, Doğrulayıcı	-Doğulayıcı yaklaşımda sadece iki faktör model yapısı kullanılmakta
flexMIRT	Ücretli	https://www.vpgcentral.com/irt-software/	-İki ve çok kategorili veriler, kayıp ve erişilmemiş veriye sahip veri setleri	-BA-EM ve MH-RM teknikleri, Maksimum olabilirlik, MAP, EAP	-Doğrulayıcı	-Açımlayıcı yaklaşımla boyutluluk analizi yapamamakta

* Yrd. Doç.Dr., Adıyaman Üniversitesi, Eğitim Fakültesi, Adıyaman-Türkiye, e-posta: gyavuz2010@gmail.com

**Doç.Dr., Hacettepe Üniversitesi, Eğitim Bilimleri Bölümü, Ankara-Türkiye, e-posta: nurid@hacettepe.edu.tr

Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, Cilt 6, Sayı 2, Kış 2015, 293-312.

Journal of Measurement and Evaluation in Education and Psychology, Vol. 6, Issue 2, Winter 2015, 293-312.

Geliş Tarihi: 09.09.2015

Kabul Tarihi: 11.20.2015

-Monte Carlo Simülasyonu						
BMIRT	Ücretsiz	http://www.bmirt.com/6271.html	- İki veya çok kategorili veri setleri	-MCMC	- Açımlayıcı, Doğrulayıcı	-MCMC tekniğinin kestirim süresi uzundur, ara yüzü yoktur, DOS komutlarıyla çalışır
DIMPACK (DIMTEST, DETECT, CCPROX/HCA)	Ücretsiz	http://psychometrictools.measuredprogress.org/dif2	- İki ve çok kategorili veriler, kayıp ve erişilmemiş veriye sahip veri setler	-Nonparametrik yaklaşımla	- Açımlayıcı, Doğrulayıcı	- Maksimum madde sayısı 150 ve maksimum örneklem büyüklüğü 7000,
Velicer'in MAP Testi ve Paralel Analiz	Ücretsiz	https://people.ok.ubc.ca/briocnn/nfactors/nfactors.html	-Çok Kategorili	-Temel bileşenler analizi	-Açımlayıcı	-Programların kullanımı için SPSS veya SAS programlarıyla syntax (betik) yazılması gerekli