

Örtük Sınıf Analiziyle Yapılan Ölçme Eşdeğerliği Çalışmalarında Model Seçimi*

Model Selection in Measurement Equivalence Studies with Latent Class Analysis

Duygu GÜNGÖR**, Mediha KORKMAZ***, Hakan Savaş SAZAK****

ÖZ: Bu çalışmanın amacı örtük sınıf analiziyle ölçme eşdeğerliğinin farklı koşullarda incelenmesidir. Örtük sınıf analizi gözlenen ve örtük değişkenlerin süresiz yapıda olduğu ya da sürekli ölçekleme düzeyinde olup normal dağılım, homojenlik, tekboyutluluk gibi varsayımların ihlal edildiği koşullarda kullanmaya uygun güçlü bir alternatiftir. Çalışma kapsamında örneklem büyüklüğü, madde sayısı, yanlılığın olduğu parametre, eşdeğer olmayan madde sayıları değiştirilerek 28 farklı koşul oluşturulmuştur. Model seçimleri BIC, CAIC, AIC ve AIC3 bilgi kriterleri kullanılarak yapılmıştır. Örneklem büyüklüğünün etkisi incelendiğinde genel olarak tüm bilgi kriterlerinin büyüklük arttığında doğru karar oranlarının arttığı gözlemlenmiştir. Madde sayısındaki değişme doğru karar oranında bir değişikliğe neden olmamıştır. Eşdeğer olmayan madde sayısı arttığında ise ölçek düzeyindeki analizlerde doğru karar oranının da arttığı söylenebilir. Yanlılığın olduğu parametrelerin hem eğim hem sabit olduğu koşullarda, sadece sabit olduğu koşullara oranla daha çok doğru karar alınmıştır. Genel olarak değerlendirildiğinde BIC ve CAIC'nin yanlış negatif, AIC'nin yanlış pozitif karar eğilimi olduğu, AIC3'ün ise daha tutarlı sonuçlar verdiği söylenebilir.

Anahtar sözcükler: örtük sınıf analizi, ölçme eşdeğerliği, bilgi kriterleri, model seçimi, simülasyon çalışması

ABSTRACT: Purpose of this study is to investigate measurement equivalence with latent class analysis in different conditions. Latent class analysis is an alternative method when both observed and latent variables are discrete, or continuous but assumptions like normal distribution, homogeneity, unidimensionality are violated. Within the study 28 different conditions were simulated by changing; sample size, number of items, inequivalence type, number of inequivalent items. BIC, CAIC, AIC and AIC3 are used for model selection. In general, for all information criteria, increasing the sample size induced the true decision rate. Number of items did not cause any alteration in true decision rates. In scale level analysis true decision rates increase parallel with the number of inequivalent items. Also when there is inequivalence both in slope and intercept parameters, more true decisions are taken with respect to the case when there is inequivalence only in intercept parameter. In conclusion BIC and CAIC tend to select false negative, AIC tends to select false positive models whereas AIC3 seems to be more consistent.

Keywords: latent class analysis, measurement equivalence, information criteria, model selection, simulation study

1. GİRİŞ

Eğitim alanında öğrenci başarısını test etme, seçme, yerleştirme gibi pek çok işlem için ölçme araçları kullanılmaktadır. Ölçme araçları psikometrik açıdan incelenirken çoğu zaman geçerlik, güvenilirlik çalışmaları yürütülmekte fakat çalışmaların pek azı farklı gruplar için geçerli ölçümler yapıp yapılmadığı sorusuna odaklanmaktadır. Söz konusu sorunun yanıtı, ölçme eşdeğerliği / ölçüm değişmezliği (measurement equivalence/invariance) kavramı altında incelenmektedir. Ölçme eşdeğerliği, bir bireyin hangi grupta yer aldığından bağımsız olarak belli bir gözlenen puana sahip olmasıdır (Horn ve McArdle, 1992; Millsap, 2011). Örnek olarak aynı başarı düzeyine sahip kadın ve erkeklerin aynı gerçek puanlara sahip olacağı düşünülebilir. Eğer

* Bu makale birinci yazarın doktora tezinin bir parçasıdır ve IV. Ulusal Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi'nde sözel bildiri olarak sunulmuştur.

** Yrd. Doç. Dr., İzmir Üniversitesi, Fen-Edebiyat Fakültesi, İzmir-Türkiye, e-posta: duygu.gungor@izmir.edu.tr

*** Doç. Dr., Ege Üniversitesi, Edebiyat Fakültesi, İzmir-Türkiye, e-posta: mediha.korkmaz@ege.edu.tr

**** Yrd. Doç. Dr., Ege Üniversitesi, Fen Fakültesi, İzmir-Türkiye, e-posta: hakan.savas.sazak@ege.edu.tr

cinsiyet farkı eşdeğer yetenek seviyesindeki kişilerde farklı gerçek puanlar gözlenmesine neden oluyorsa ölçme eşdeğerliğinden söz edilemez. Alanyazındaki ölçme eşdeğerliği çalışmalarının çoğu yapısal eşitlik modellemeleri ya da madde tepki kuramı kapsamındaki yöntemlerle yapılmaktadır (Somer, Korkmaz, Dural ve Can, 2009). Yeni bir yöntem olarak ise örtük sınıf analizleriyle (ÖSA) ölçme eşdeğerliği çalışmaları dikkat çekmektedir (örn; Eid, Langeheine ve Diener, 2003; Güngör, Korkmaz ve Somer, 2013; Moors ve Wennekers, 2003; Kankaras, Moors ve Vermunt, 2010).

Bu çalışmanın amacı, Monte Carlo simülasyonu kullanılarak, ölçekteki toplam ve yanlış madde sayısı, örneklem büyüklüğü, yanlışlık türü ve model seçim sürecinde kullanılan bilgi kriterine bağlı olarak örtük sınıf modellerinin ölçme eşdeğerliğini madde ve ölçek düzeyinde tahmin etme gücünü incelemektir. Bu kısımda ÖSA ile ölçme eşdeğerliğinin test etme ve model seçim sürecine değinilecek, sonrasında yapılan simülasyon çalışması aktarılacaktır.

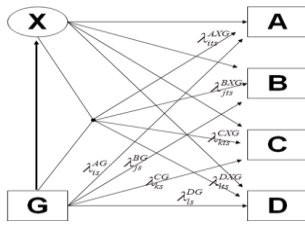
1.1. Örtük Sınıf Analiziyle Ölçme Eşdeğerliğinin İncelenmesi

Eğitimde ve psikolojide araştırmaların büyük bir kısmında, gözlenen değişkenlerden, örtük (gözlenemeyen) yapılar incelenmektedir. Geleneksel yöntemler kullanıldığında, değişkenin süresiz yapıda olması, normal dağılım göstermemesi, homojen olmaması metodolojik problemlerle yol açmaktadır. Örtük sınıf modelleri için kullanılan beklenti-maksimizasyon algoritmasının formüle edilmesi ve bilgisayar yazılımlarında kullanılmaya başlanmasıyla, son yıllarda örtük sınıf analiziyle yapılan çalışma sayısı oldukça artmıştır.

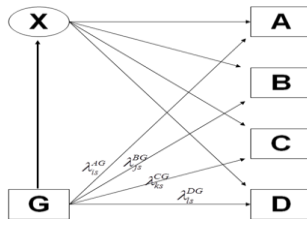
ÖSA en genel haliyle hem gözlenen hem örtük değişkenin süresiz yapıda olduğu durumlarda kullanılan bir örtük değişken modeli olarak tanımlanabilir. Örtük sınıf analizinde koşullu olasılıklar ve örtük sınıf olasılıkları olmak üzere iki temel parametre tahminlenmektedir. Koşullu olasılıklar faktör analizindeki faktör yüklerine benzemekte ve örtük yapının özelliği ile ilgili bilgi vermektedir. Örtük sınıf olasılıkları ise farklı sınıfların büyüklüğünü/yaygınlığını göstermektedir. Örtük sınıf modellerinin kullanıldığı alanlara ve çalışmalara örnek olarak psikoloji (Bornovalova, Levy, Gratz ve Lejuez, 2010; Campbell, Morgan-Lopez, Cox ve McLoyd, 2009; Klonsky ve Olino, 2008) ve eğitim (Brown 2007; Güngör-Culha ve Korkmaz, 2011; Waasdorp ve Bradshaw, 2011) verilebilir.

ÖSA ile ölçme eşdeğerliği çalışmaları ise oldukça yenidir. Örtük sınıf modelleri ile yapılan ölçme eşdeğerliği çalışmalarında da diğer yöntemlerle benzer olarak temel amaç eşdeğerlik seviyesi en yüksek olan modeli tespit etmektir. Model seçim süreci genellikle karşılaştırma grupları için gerekli sayıda örtük sınıfın ya da örtük faktörün tespit edilmesi ile başlamakta ve iç içe geçmiş çeşitli modeller test edilmektedir. Aşağıda heterojen modelden başlayarak tamamen homojen modele doğru hiyerarşik bir sırayla modeller açıklanmıştır. Uygulamada en üst (tamamen homojen) ya da en alt (heterojen) seviyedeki modellerin herhangi birinin ilk model olarak kullanıldığı araştırmalara rastlanabilmektedir. Örtük sınıf modelleri ile ölçme eşdeğerliği çalışmaları oldukça yeni olduğu için alan yazında modellere heterojen, kısmi homojen ve homojen modeller dendiği gibi, yapısal eşitlik modelleri kapsamındaki terimlerle de karşılaşılmaktadır (örn: Morren, Gelissen ve Vermunt, 2012).

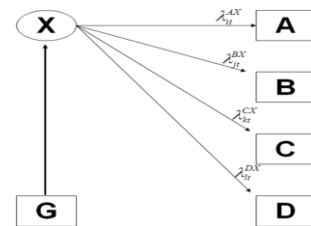
G grup, X örtük, A, B, C ve de gözlenen değişkenler olmak üzere modeller şekildeki gibidir.



Şekil 1.1. Heterojen Model



Şekil 1.2 Kısmi Homojen Model



Şekil 1.3. Yapısal Homojenlik Modeli

1.1.1. Heterojen Model

Heterojen model tahmin edilecek parametrelere (koşullu olasılıklar ve örtük sınıf olasılıkları) hiçbir kısıt konmamış; diğer bir ifadeyle parametrelerin karşılaştırma grupları için ayrı ayrı tahminlenmesine izin verildiği durumu göstermektedir (McCutcheon ve Hagenaars, 1997). Heterojen modelde tüm model parametreleri yani olasılıksal terimlerle örtük sınıf olasılıkları ve koşullu olasılıklar, loglineer terimlerle direkt etki ve ortak etki, lojistik terimlerle de sabit ve eğim parametreleri gruba özgüdür. Bu koşulda ölçme eşdeğerliği yoktur ve gruplar arası karşılaştırma yapmak mümkün değildir.

1.1.2. Kısmi Homojen Model

Parametrelerin bir kısmına kısıt konarak test edilen modeller kısmi homojen model olarak adlandırılmaktadır. Grubun belli sayıda gözlenen değişken üzerindeki direkt etkisi sınırlandırılarak, grup örtük değişken ortak etkisini sadece iki madde için sınırlandırarak, gözlenen değişkenlerin hata miktarlarını eşitleyerek vb. çeşitli kısmi homojen modeller test edilebilmekle birlikte, grup örtük değişken ortak etkisinin modelden kaldırıldığı kısmi homojen model özellikle önemlidir (Kankaras, Moors ve Vermunt, 2010). Kısmi homojen modelde grup örtük değişken ortak etkisi modelden kaldırılmıştır ve bu modelin kabulüyle örtük sınıf olasılıklarının gruplar arasında farklılaşıp farklılaşmadığı incelenebilir (Kankaras, Moors ve Vermunt, 2010). Lojistik terimlerle bu durum eğim parametrelerinin gruplar için eşitlenmesi şeklinde tanımlanabilir. Bu modelde karşılaştırma gruplarının gözlenen değişkenler üzerindeki direkt etkisi başka bir deyişle gruba özgü sabit parametreleri serbesttir. Yapısal eşitlik modellerindeki metrik eşdeğerliğe, madde tepki kuramındaki düzgün formulu madde işlevsel farklılığına benzeyen bu modelde; gözlenen değişken üzerinde grup örtük değişken ortak etkisi yoktur (Şekil 1.2). Başka bir deyişle eğim parametreleri karşılaştırma grupları için eşittir. Bu da gözlenen ve örtük değişken arasındaki ilişkinin karşılaştırma grupları için eşit olduğu anlamına gelmektedir. (Kankaras, Moors ve Vermunt, 2010; McCutcheon ve Hagenaars, 1997).

1.1.3. Yapısal Homojenlik Modeli

Kısmi homojenlik modelinin bir sonraki aşaması ve karşılaştırma grupları arasındaki farkın güvenilir şekilde incelenebileceği aşama yapısal homojenliktir. Bu basamak homojen model olarak da adlandırılmaktadır (Kankaras ve Moors, 2009). Yapısal eşitlik modellerindeki skalar (güçlü) eşdeğerlik modeli ve madde tepki kuramındaki madde güçlük ve ayırt edicilik parametrelerinin gruplararası eşdeğer olduğu modellere benzetilebilir. Homojen modelde karşılaştırma gruplarının gözlenen değişkenler üzerindeki direkt etkisi de modelden kaldırılmıştır ve örtük değişken üzerinden tanımlanmaktadır. Bu koşul loglineer terimlerle grup örtük değişken ortak etkisine ek olarak grubun direkt etkisi de 0'a eşitlenerek, lojistik terimlerle de eğim parametresi ve sabit parametreleri gruplar için eşitlenerek sağlanabilir (Şekil 1.3). Pratikte eşdeğerliğin sağlanması için yapısal homojen model yeterli sayılmaktadır. Ölçme eşdeğerliği sağlandığında grupların örtük yapıda gösterdikleri dağılım farklılaşabilir, fakat karşılaştırma gruplarının örtük yapı ile ilişkisi eşdeğerdir (Collins ve Lanza, 2010).

Buraya kadar ölçme eşdeğerliği ölçek düzeyinde, yani gözlenen değişkenlerin tümü için aynı anda kısıt konan modeller açıklanmakla birlikte, örtük sınıf modelleri madde düzeyinde de eşdeğerlik çalışması yapmak için uygundur. Bir ölçek üzerinden düşünüldüğünde maddelerin hepsi birden eşdeğer olmayabilir. Performans testlerinde bazı maddeler içerikleri itibarıyla karşılaştırma gruplarından biri için daha kolay olabilir, ya da bir tutum ölçeğinde toplumsal normlar karşılaştırma gruplarından birinin herhangi bir maddeye daha çok “tamamen katılıyorum” demesine neden olacak yanlılıkta olabilir. Gözlenen değişkenlerin davranışlar olduğu durumlarda da örneğin risk alma davranışının “hızlı araba kullanma” gibi bir göstergesi yaşa bağlı olarak değişirken, diğer göstergeler olabilecek “zararlı madde kullanımı, hırsızlık” gibi davranışlar karşılaştırma gruplarında aynı koşullu olasılıklara sahip olabilirler. Böylesi durumların tespiti için tüm gözlenen değişkenlere aynı anda kısıt koymaya alternatif olarak madde düzeyinde eşdeğerlik çalışmaları yapılmaktadır. Genellikle karşılaştırma gruplarının örtük sınıf ya da faktör sayısı belirlendikten sonra, homojen model test edilmektedir. Homojen modelde gösterge değişkenler için artık değerler (residual) hesaplanıp göz önünde bulundurularak, sırasıyla artık değeri en çok olan maddeden başlanarak, diğer maddelere de grubun direkt etkisi serbest bırakılarak iç içe geçmiş kısmi homojen modeller test edilmektedir. Tüm maddelerde grubun direkt etkisine kısıt konulan model bir önceki bölümdeki kısmi homojen modelin aynısıdır. Temel model olarak heterojen modelin alındığı durumlarda ise bir sonraki basamak olarak grup örtük değişken ortak etkisine kısıt konmaktadır. Bir üst model olarak grup direkt etkisine konulan kısıtlar maddeler için teker teker eklenmektedir.

Alanyazında ÖSA ile ölçme eşdeğerliği çalışmalarına Eid, Langeheine ve Diener (2003), Moors ve Wennekers (2003), Kankaras ve Moors (2009), Güngör, Korkmaz ve Somer (2013) verilebilir. Örneğin; Eid, Langeheine ve Diener (2003), üniversite öğrencilerinin yaşam doyumlarının incelendiği uluslararası bir araştırmadan, katılımcıların ekonomik durum, arkadaşlık ilişkileri, aşk hayatları ve kendilerinden ne kadar memnun oldukları sorularına verdikleri yanıtları çoklu grup örtük sınıf modeli ile incelemişlerdir. Cevaplar 7’li Likert tipindedir (1= hiç memnun değilim, 7= tamamen memnunum), fakat araştırmada yorumlama kolaylığı açısından 3 kategoride toplanarak analiz edilmiştir. Otuzdokuz ülkenin var olduğu çalışmadan Çin (n=522) ve Birleşik Devletler (n=432)’den elde edilen veriler alınarak aralarındaki eşdeğerlik test edilmiştir. Araştırmacılar örtük sınıf analizinin kullanılma nedenleri, model seçim süreçlerini detaylı şekilde açıkladıkları araştırmanın analiz kısmında, öncelikle tüm parametrelerin ülkelere özgü olduğu (heterojen) modeli test etmişlerdir. Sonrasında tespit edilen 2 sınıflı modeli, tüm koşullu olasılıkların eşdeğer sayıldığı (homojen) modeli de ekleyerek karşılaştırmışlar ve ölçme eşdeğerliğini red etmişlerdir. Bu varsayımı yumuşatmak için sadece bir örtük sınıfta koşullu olasılıkların eşdeğer diğerinde ise farklı varsayıldığı kısmi homojen modeli test ederek, en iyi uyum sağlayan modelin kısmi homojen model olduğunu bulmuşlardır.

1.2. Model Seçimi

Örtük sınıf analizi model seçim sürecinde, veriye uyum sağlayan en sade (parsimony) yani en az sayıda örtük sınıfa sahip ve daha az parametre tahminlenen model aranmaktadır. Model seçim sürecinde istatistiksel kriterler, sadelik ve yorumlanabilirlik göz önüne alınmalıdır (Collins ve Lanza, 2010; Silvia, Kaufman ve Pretz, 2009). Bu koşulları sağlayan modeli bulmak için; sınıflandırma kalitesi ile ilgili bilgi veren entropi (entropy), temelinde gözlenen değerler ile beklenen değerlerin karşılaştırılması yatan Pearson ki-kare (χ^2), en çok olabilirlik ki-kare oranı (likelihood ratio chi square) (L^2)¹ gibi istatistikler ve Akaike bilgi kriteri (Akaike Information Criteria) (AIC), Bayesian bilgi kriteri (Bayesian Information Criteri) (BIC) gibi bilgi kriterleri kullanılabilmektedir (McCutcheon, 2002; Silvia, Kaufman ve Pretz, 2009). Örtük sınıf

¹ Bazı kaynaklarda G^2 olarak da gösterilmektedir. Örneğin (Collins ve Lanza, 2010; Uebersax, 1994)

modellerinde sınıf sayısı belirlenirken karşılaştırılmalı olarak seçim yapılmaktadır, amaç doğru modeli bulmaktan çok, daha fazla bilgi sağlayan modeli tespit etmektir (Moors ve Wennekers, 2003).

Model seçim sürecinde genel olarak analizlere gözlenen değişkenlerin bağımsızlığını ifade eden sınıf sayısının 1'e eşit olduğu modelle başlanır (Magidson ve Vermunt, 2004). Eğer bu model kabul edilirse örtük sınıf analizine ihtiyaç kalmayacaktır. Ancak bu temel model verilere uyum sağlamazsa yani en az iki örtük sınıf tanımlanabilecekse analizler her seferinde sınıf sayısı bir arttırılarak serbestlik derecesinin izin verdiği sayıya kadar devam ettirilebilir. En iyi modeli seçmek için kullanılan yöntemlerden biri L^2 fark istatistiklerine göre karar vermektir. Karşılaştırılacak iki model için hesaplanan L^2 değerleri arasındaki fark hesaplanarak ΔL^2 değeri elde edilmektedir. ΔL^2 'nin istatistiksel olarak anlamlı olup olmadığı da yine söz konusu iki modelin serbestlik dereceleri arasındaki farkın kritik değeriyle karşılaştırılması ile elde edilir. Model uyumunu L^2 fark istatistikleri ile incelemek özellikle dağılımın belirsiz olduğu durumlarda sorunlu olabilmektedir (Dias, 2006; Morren, Gelissen, ve Vermunt, 2012; Yang ve Yang, 2007). Çok sayıda koşullu olasılık tahminlenmesi gereken, serbestlik derecesinin yüksek olduğu koşullarda da L^2 fark istatistikleri genellikle χ^2 dağılımı göstermez (Collins ve Lanza, 2010). Değişken sayısının fazla olduğu durumlarda karşılaşılan ve L^2 istatistiklerini güvenilir kılan bir diğer sorun da bazı hücrelerin boş kalma olasılığıdır.

Alternatif bir yol olarak AIC, BIC bilgi kriterlerinin kullanılması genel olarak kabul görmüştür. Bununla birlikte CAIC, AIC3, DBIC gibi türevleri de var olan kriterlerden hangisinin kullanılmasının uygun olduğuna dair bir görüş birliği bulunmamakta, alan yazında bu konu ile ilgili pek çok yayın bulunmaktadır (Bauer ve Curran, 2003; Dias, 2006; Lin, 2006; 2012; Nylund, Asparouhov, ve Muthen, 2007; Vrieze, 2012; Yang ve Yang, 2007). Farklı simülasyon koşullarında Dias (2006) AIC3, Nylund, Asparouhov, ve Muthen (2007) BIC'nin daha iyi sonuçlar verdiğini raporlamışlardır. Kankaras, Vermunt, ve Moors (2011) ölçme eşdeğerliğinin farklı yöntemler ile (çoklu grup doğrulayıcı faktör analizi, madde cevap kuramı ve örtük sınıf faktör analizi) karşılaştırıldıkları simülasyon çalışmasında, madde düzeyinde örneklem sayısı küçük olduğunda AIC'nin, örneklem sayısı arttığında ise BIC'nin AIC'den daha iyi sonuçlar verdiğini tespit etmişlerdir. Ölçek düzeyinde ise AIC'nin daha iyi sonuçlar verdiğini raporlanmıştır. Morren, Gelissen ve Vermunt (2011; 2012) örtük sınıf faktör analizi kullanarak yaptıkları ölçme eşdeğerliği çalışmalarında BIC değerini kullanmayı tercih etmişlerdir.

Bilgi kriterleri kullanılarak yapılan model seçim süreci diğer modeller arasından en iyisini seçmeye yöneliktir ve hangi bilgi kriteri kullanılırsa kullanılsın, kriterin aldığı daha düşük değer, daha iyi model anlamına gelmektedir.

2. YÖNTEM

2.1. Araştırma Deseni

Araştırma deseninde, toplam madde sayısı (5 ve 10 madde), eşdeğer olmayan madde sayısı (0,1,3, 5 ya da 10), örneklem büyüklüğü (100 ve 500), yanlılık türü (hem sabit hem eğim parametresinde, sabit parametresinde) ve model seçiminde kullanılan bilgi kriteri (BIC, CAIC, AIC, AIC3) olmak üzere beş bağımsız değişken yer almaktadır. Model seçiminde kullanılan bilgi kriterleri veri setlerinin üretilmesi aşamasında değil; analiz sonuçlarının yorumlanması aşamasında devreye girmektedir. Bu kapsamda 28 farklı koşul ve her koşul için 100 tekrar; toplamda 2800 veri seti üretilmiş ve analiz edilmiştir.

Simülasyon koşulları oluşturulurken örtük sınıf analizi ile üç farklı modelin de tespit edilme gücünü araştırabilmek için, farklı homojenlik seviyelerinde (heterojen, kısmi homojen,

homojen)² veri setleri üretilmiştir. Böylelikle örtük sınıf analizi kullanılarak yapılan ölçme eşdeğerliği çalışmalarında sadece yanlış pozitif ve sadece yanlış negatif karar oranları ve iki hatalı karar türünün birlikte incelenmesi mümkün olmuştur. Örneğin, veri setinin heterojen modelle uyumlu olduğu koşullarda, kısmi homojen ya da homojen modelin daha iyi model olarak seçilmesi, özetle eşdeğerlik yokken, var kararı verilmesi yanlış negatif kararı temsil etmektedir. Benzer şekilde veri seti homojen modelle uyumluyken - eşdeğerlik söz konusuysa - kısmi homojen ya da heterojen modelin seçilmesi de yanlış pozitif kararın temsil edicisidir (Kankaras, Vermunt, ve Moors, 2011). Kısmi homojen modelin temsil edildiği koşullarda ise doğru, yanlış negatif ve yanlış pozitif olmak üzere üç farklı karar çıktısı elde etmek mümkündür.

Farklı homojenlik seviyelerine uygun veri setlerinin üretilmesi aşamasında eşdeğer olmayan madde sayısı ve yanlışlık türü değişkenlerinden yararlanılmıştır. Hem eğim, hem sabit parametrelerinde yanlışlık olan koşullar heterojen modellerin, sadece sabit parametresinde yanlışlık olan koşullar ise kısmi homojen modellerin üretilmesinde kullanılmıştır. Her iki parametrenin eşdeğer olduğu, dolayısıyla eşdeğer olmayan madde sayısının sıfır olduğu koşullar ise homojen modellerin temsilcisidir.

Araştırma kapsamında sabitlenen koşullar ise; grup sayısı (2), gruplar için örneklem büyüklüğü (100'er ya da 500'er), örtük sınıf sayısı (2), örtük sınıf olasılıkları (.50), madde kategori sayısı (5), eşdeğer olmayan maddelerin ölçekteki sırası (3; 3-4-5) şeklindedir.

2.2. Araştırmada Kullanılan İstatistik Paket Programı

Araştırmanın hem veri setlerinin üretilmesi, hem de analizleri Latent Gold 4.5 Advanced (Vermunt & Magidson, 2005) programının eklentisi olan LG-Syntax Module (Vermunt ve Magidson, 2008) programı kullanılarak yürütülmüştür. Bu eklenti ile Monte Carlo simülasyon çalışmaları yapmak, analiz edilen modellere çok çeşitli kısıtlar koymak, lojistik ya da loglineer formlarda ölçme eşdeğerliği çalışmaları yapmak mümkündür. Latent Gold programı parametre tahminlerini EM ve Newton-Raphson algoritmalarının karmasını kullanarak yapmaktadır. Latent GOLD programı kullanıcı dostu yazılımıyla ölçme eşdeğerli çalışmalarında sıklıkla kullanılmaya başlanmıştır. Program L^2 ya da LL'ye dayalı olarak bu çalışmada da kullanılan AIC, BIC, AIC3 ve CAIC bilgi kriterlerinin değerlerini vermektedir.

2.3 Simülasyon Verisinin Özellikleri

2.3.1 Eğim ve Sabit Parametrelerinin Değişimlendiği Koşullar

Grup karşılaştırmalarının mümkün olmadığı heterojen modeli temsil eden veri setleri hem eğim hem sabit parametreleri değişimlenerek üretilmiştir. Bu kapsamda 1. grup için belirlenen sabit değerleri "0, -1, 0.5, 1, 0.5", eğim değeri de "0.5" şeklindedir. 2. grup için aynı değerler ise sırasıyla "0, -2, -1, 0, 1" ve eğim değeri "1.5" olarak atanmıştır³. Maddelerden sadece sabit parametreleri farklı, eğim parametreleri eşdeğer olan maddeler için ise eğim parametresine verilen değer "1" olarak belirlenmiştir. Eğim parametresindeki yanlışlıklar, örtük değişkenle ilişkinin gücünü etkilemekte, bu durum örtük sınıf olasılıklarında artış olarak kendini göstermektedir (Kankaras, Vermunt, ve Moors, 2011).

² Sadece bazı parametrelere sınır konmuş çeşitli kısmi homojen modeller oluşturulabilmekle birlikte bu çalışmada tüm eğim parametrelerinin eşdeğer olduğu modeller kısmi homojen model olarak incelenmiştir.

³ Katsayılar arasındaki farklar belirlenirken Kankaras, Vermunt, ve Moors, 2011 tarafından yapılan simülasyon çalışması örnek alınmış, böylece orta derecede zorlukla tespit edilecek koşullar oluşturulması sağlanmıştır. Bimodal dağılım elde edilmesini sağlayan katsayılar kullanıldığında elde edilen koşullu olasılık değerleri Ek.1'de sunulmuştur.

2.3.2. Sabit Parametresinin Değişimlendiği Koşullar

Yanlılığın olduğu parametrenin sabit parametresi olduğu koşullar için üretilen veri setlerinde, hem eğim hem sabit parametresinin değişimlendiği koşullarla aynı şekilde, eşdeğer olan maddeler için sabit değerleri her iki grup için de; “0, -1, 0.5, 1, 0.5” eğim değerleri de “1” olarak belirlenmiştir. Eşdeğer olmayan madde olarak değişimlenen maddelerde ise 1. grup için diğer maddelerle aynı sabit ve eğim parametreleri kullanılmıştır. 2. grup için ise kategorilere göre sırasıyla sabit değerleri “0, -2, -1, 0, 1”, eğim parametresi ise 1. grupla aynı şekilde “1” olarak atanmıştır.

2.3.2 Maddelerin Tamamının Eşdeğer Olduğu Koşullar

Maddelerin tamamının eşdeğer olduğu koşullarda da yukarıdaki iki durumla aynı katsayılar kullanılmıştır. Bir başka deyişle, hem grup 1 hem de grup 2 için sabit değerleri “0, -1, 0.5, 1, 0.5” eğim değerleri de “1” ‘dir.

Yukarıdaki değerler kullanılarak elde edilen veri setlerine ait koşullu olasılık değerleri Ek.1’de sunulmuştur.

2.4 İşlem Yolu

Çalışma kapsamında beş aşamadan oluşan bir işlem yolu izlenmiştir. Yirmisekiz farklı koşul için veri setleri Latent Gold 4.5 (Vermunt ve Magidson, 2005) programı kullanılarak üretilmiştir. Elde edilen veri setlerinin ölçek ve madde düzeyinde ölçme eşdeğerliği analizleri yapılmıştır. Analizler sonucunda elde edilen L^2 değerleri kullanılarak, her koşul ve analiz için dört farklı bilgi kriterine (BIC, CAIC, AIC, AIC3) ait değerler hesaplanmıştır. Elde edilen bilgi kriteri değerleri kullanılarak model seçimi yapılmıştır. Seçilen modellere ilişkin doğru karar yüzdeleri tespit edilerek raporlanmıştır.

3. BULGULAR

Bulgular örneklem büyüklüğü, eşdeğer olmayan madde sayısı ve yanlılığın olduğu parametrelere göre ölçek ve madde düzeyinde verilen kararların ortalama ve standart sapmaları hesaplanarak sunulmuştur. Sunulan ortalama değerlerinde doğru karar için 28 koşul varken, örneğin heterojen model söz konusu olduğunda yanlış negatif karar ihtimali bulunmamakta ve daha az koşul sayısı bulunmaktadır. Bu nedenle ortalamalar sütun bazında yorumlanmakta, satır toplamaları 100 olmamaktadır.

3.1. Örneklem Büyüklüğünün Model Seçimine Etkisi

Örneklem büyüklüğünün model seçimine etkisini incelemek amacıyla farklı koşullardaki karar ortalamaları ve standart sapmaları hesaplanmış ve Tablo 1’de sunulmuştur. Her iki örneklem büyüklüğünde de 14 farklı koşulda elde edilen sonuçların ortalamaları; ölçek düzeyinde doğru karar, yanlış pozitif ve yanlış negatif kararlar olarak, madde düzeyinde ise eşdeğer olmayan maddelerin yanlılığın tespit edildiği doğru karar ortalamaları, eşdeğer olan maddelerde de yanlılık tespit edilen yanlış pozitif karar ortalamaları sunulmuştur.

Ölçek düzeyindeki analizler incelendiğinde BIC ve CAIC kullanıldığında beklenen şekilde örneklem büyüklüğüne paralel olarak doğru karar oranları artmakta, yanlış negatif karar oranları azalmaktadır. Örneklem büyüklüğü 500’e yükseldiğinde de BIC ile ortalama %23, CAIC ile ortalama %30 yanlış negatif karar oranı, yanlılığı tespit etmek için yeterli olmadığını göstermektedir. Madde düzeyinde ise örneklem büyüklüğü 500’e yükseldiğinde doğru şekilde yanlılık tespit edilebilmiştir. AIC3 kullanılarak yapılan model seçimlerinde örneklem büyüklüğü 500’e yükseldiğinde yanlış negatif oranı oldukça düşüktür. Ölçek düzeyindeki analizler

incelendiğinde tüm bilgi kriterleri içinde en düşük yanlış negatif oranı ile AIC yanlılığı en iyi şekilde tespit eden bilgi kriteri olarak öne çıkmaktadır. Madde düzeyinde bakıldığında ise en yüksek yanlış pozitif oranının da yine AIC'ye ait olduğu görülmektedir.

Tablo 1: Örneklem Büyüklüğüne Göre Model Seçim Ortalama ve Standart Sapmaları

ÖRNEKLEM BÜYÜKLÜĞÜ	BİLGİ KRİTERİ	ÖLÇEK DÜZEYİNDE						MADDE DÜZEYİNDE			
		Doğru Karar		Yanlış Pozitif		Yanlış Negatif		Doğru Karar ^a		Yanlış Pozitif ^b	
		Ort.	S.s.	Ort.	S.s.	Ort.	S.s.	Ort.	S.s.	Ort.	S.s.
100	BIC	43.57	45.40	0.00	0.00	65.83	41.97	54.96	10.51	0.39	0.65
	CAIC	35.50	41.64	0.00	0.00	75.25	39.86	40.04	12.16	0.15	0.40
	AIC	81.71	29.33	3.13	2.75	19.25	31.11	95.87	2.68	13.35	3.89
	AIC3	67.79	42.59	0.25	0.46	37.42	43.97	87.63	4.75	4.13	2.24
500	BIC	80.35	36.50	0.00	0.00	22.92	38.64	99.98	0.00	0.01	0.09
	CAIC	74.21	41.64	0.00	0.00	30.08	43.69	99.96	0.00	0.01	0.09
	AIC	97.86	3.01	3.63	3.07	0.08	0.29	100	0.00	12.93	3.64
	AIC3	96.86	10.91	0.25	0.46	3.50	11.81	100	0.00	4.20	2.49

^a Sadece eşdeğer olmayan maddelere ilişkin doğru sonuçlardır.

^b Sadece eşdeğer maddelere ilişkin yanlış pozitif sonuçlardır.

3.2. Eşdeğer Olmayan Madde Sayısının Model Seçimine Etkisi

Eşdeğer olmayan madde sayısının model seçimine etkisini incelemek için, yanlılığın olduğu parametrenin hem eğim hem sabit ve sadece sabit olduğu parametrelerde yapılan seçimlerin farklı örneklem büyüklüklerindeki ortalama ve standart sapmaları hesaplanmıştır (Tablo 2). Örneklem büyüklüğünün 100 olduğu koşullarda ölçek düzeyinde yapılan analizlerde BIC, CAIC kullanıldığında 10 maddeden bir ve üç, 5 maddeden bir yanlış madde varlığında doğru karar oranı 0'a yakinken (< 6); yanlış negatif karar oranları da ortalama %60'ın üzerindedir. AIC3 kullanıldığında eşdeğer olmayan madde oranı yükseldikçe doğru karar oranı da yükselmiştir (sırasıyla; 6.6, 28.50, 51.5, 88.5). Ölçek düzeyinde en yüksek doğru karar ortalaması ve en düşük yanlış negatif karar ortalaması AIC ile yapılan seçimlere ait olmakla birlikte madde düzeyinde bakıldığında ise en yüksek yanlış pozitif karar ortalamasının AIC'ye ait olduğu görülmektedir. Örneklem büyüklüğü 500'e yükseldiğinde tüm bilgi kriterlerinin daha iyi sonuçlar verdiği görülmektedir. Özellikle madde düzeyindeki analizlerde doğru karar ortalamaları %100'dür. AIC ile yapılan kararlarda ise madde düzeyinde yanlış pozitif karar ortalaması %10'un üzerindedir. Ölçek düzeyindeki analizlerde BIC ve CAIC ile alınan kararlarda eşdeğer olmayan madde oranı yükseldikçe doğru karar ortalaması yükselmiştir. AIC3 sadece 10 maddeden biri yanlışken daha düşük doğru karar ortalamasına sahipken, bu oran diğer üç koşulda %97'nin üzerindedir. AIC kullanıldığında ise eşdeğer olmayan madde oranından bağımsız olarak doğru kararlar alınmıştır. Ancak yanlış pozitif karar da yine sadece AIC kullanıldığında ortaya çıkmaktadır.

Tablo 2: 100 ve 500 Örneklem Büyüklüğünde Eşdeğer Olmayan Madde Oranına Göre Model Seçim Yüzde Ortalama ve Standart Sapmaları

	EŞDEĞER OLMAYAN MADDE SAYISI	BİLGİ KRİTERİ	ÖLÇEK DÜZEYİNDE				MADDE DÜZEYİNDE				
			Doğru Karar		Yanlış Pozitif _a	Yanlış Negatif		Doğru Karar _b		Yanlış Pozitif _c	
			Ort.	S.s.	Frekans	Ort.	S.s.	Ort.	S.s.	Ort.	S.s.
Örneklem Büyüklüğü = 100	1/10	BIC	0.00	0.00	0	66.67	56.87	55.5	16.26	0.11	0.32
		CAIC	0.00	0.00	0	66.67	33.01	41	16.97	0.00	0.00
		AIC	28	33.94	0	48	48	97	1.41	13.39	4.59
		AIC3	6.6	9.19	0	62.33	54.37	89	4.24	2.89	1.93
	1/5	BIC	1	1.41	0	66	55.47	58	19.79	0.13	0.35
		CAIC	0.5	0.71	0	66.33	48.29	46	24.04	0.00	0.00
		AIC	66	35.36	5	21	28.93	97.5	0.7	14.88	4.02
		AIC3	28.50	36.06	0	47.67	48.52	92.5	0.7	4.5	2.27
	3/10	BIC	6	8.49	0	62.67	54.6	58.67	12.16	0.57	0.65
		CAIC	0.5	0.71	0	66.33	57.45	45.33	14.31	0.5	0.52
		AIC	87	18.39	2	8	13.86	96.67	1.75	12.14	4.66
		AIC3	51.5	61.52	0	32.33	51.73	90	4.77	3.79	2.46
	3/5	BIC	44	59.4	0	37.33	53	57.67	12.11	0.00	0.00
		CAIC	34.5	48.79	0	43.67	51.19	44.33	14.94	0.00	0.00
		AIC	97.5	3.54	5	0.00	0.00	97	1.79	15	5.09
		AIC3	88.5	16.26	1	7.33	12.7	87.67	1.75	3	1.63
Örneklem Büyüklüğü = 500	1/10	BIC	20.50	28.99	0	53	50.27	100	0.00	0.00	0.00
		CAIC	8.50	12.02	0	61	53.51	100	0.00	0.00	0.00
		AIC	97	4.24	5	0.34	0.58	100	0.00	13.72	3.59
		AIC3	79.50	28.99	0	13.67	23.67	100	0.00	3.66	2.52
	1/5	BIC	52	65.05	0	32	53.70	100	0.00	0.00	0.00
		CAIC	46	65.05	0	36	55.57	100	0.00	0.00	0.00
		AIC	97.5	3.53	5	0.00	0.00	100	0.00	11.13	2.9
		AIC3	99	1.41	1	0.33	0.58	100	0.00	2.75	2.18
	3/10	BIC	90	14.14	0	6.67	11.55	100	0.00	0.00	0.00
		CAIC	65	49.5	0	23.33	40.42	100	0.00	0.00	0.00
		AIC	99	1.41	2	0.00	0.00	100	0.00	12.5	2.88
		AIC3	100	0.00	0	0.00	0.00	100	0.00	3.5	2.14
	3/5	BIC	100	0.00	0	0.00	0.00	100	0.00	0.00	0.00
		CAIC	100	0.00	0	0.00	0.00	100	0.00	0.00	0.00
		AIC	96	5.66	8	0.00	0.00	100	0.00	10.5	3.32
		AIC3	100	0.00	0	0.00	0.00	100	0.00	2.5	2.38

^a Sadece kısmi homojen modellere ait yanlış pozitif karar olasılığı var olduğundan, 1 koşul vardır; sunulan değer bu koşula ait yanlış pozitif karar frekansıdır.

^b Sadece eşdeğer olmayan modellere ilişkin doğru sonuçlardır.

^c Sadece eşdeğer modellere ilişkin yanlış pozitif sonuçlardır.

3.3. Yanlılığın Olduğu Parametrelerin Model Seçimine Etkisi

Yanlılığın olduğu parametrelerin model seçimine etkisi, hem eğim hem sabit ve sadece sabit parametrelerinin yanlı olduğu koşullardaki doğru, yanlış pozitif ve yanlış negatif karar ortalama ve standart sapmaları üzerinden değerlendirilmiştir. Tablo 3'te de görüldüğü üzere ölçek düzeyindeki analizlerde eğim ve sabit parametrelerinin yanlı olduğu yani doğru modelin heterojen model olduğu koşullarda yanlış pozitif karar ihtimali bulunmamaktadır. Bu koşullar için madde düzeyindeki doğru karar oranları sadece eğim parametresinin sabitlendiği, yanlış pozitif karar oranları ise hem eğim hem sabit parametresinin serbest bırakıldığı modeller üzerinden sunulmuştur. Sabit parametresinin yanlı olduğu koşullara ilişkin madde düzeyindeki

analiz sonuçlarında ise benzer şekilde sadece sabit parametresinin eşitlendiği (doğru karar) ya da serbest bırakıldığı (yanlış pozitif) modellere ilişkin bulgular mevcuttur.

Tablo 3: 100 ve 500 Örneklem Büyüklüğünde Yanlılığın Olduğu Parametrelere Göre Model Seçim Yüzde Ortalama ve Standart Sapmaları

Yanlılığın Olduğu Parametre	BİLGİ KRİTERİ	ÖLÇEK DÜZEYİNDE						MADDE DÜZEYİNDE			
		Doğru Karar		Yanlış Pozitif		Yanlış Negatif		Doğru Karar ^a		Yanlış Pozitif ^b	
		Ort.	S.s.	Ort.	S.s.	Ort.	S.s.	Ort.	S.s.	Ort.	S.s.
Eğim+ Sabit n=100	BIC	50	50.09	-	-	25	42.06	61.30	10.79	0.45	0.59
	CAIC	44.83	49.61	-	-	27.58	36.94	48.96	11.16	0.31	0.47
	AIC	90.5	19.2	-	-	4.75	13.86	95	3.3	16.86	4.96
	AIC3	77	36.16	-	-	11.5	27.17	87.57	6.03	4.95	1.99
Eğim+Sabit n=500	BIC	89.83	23.94	-	-	5.08	16.99	99.96	0.21	0.00	0.00
	CAIC	84.83	33.39	-	-	7.58	23.86	99.91	0.29	0.00	0.00
	AIC	100	0.00	-	-	0.00	0.00	100	0.00	14.68	2.61
	AIC3	100	0.00	-	-	0.00	0.00	100	0.00	4.95	1.89
Sabit n=100	BIC	18.33	27.65	0.00	0.00	81.67	27.65	48.52	4.79	0.05	0.21
	CAIC	4.66	8.16	0.00	0.00	95.33	8.16	31.13	3.41	0	0
	AIC	67.33	37.58	3.67	2.88	29	39.15	96.74	1.48	9.96	2.73
	AIC3	47.83	49.14	0.33	0.52	51.83	49.47	87.69	3.11	2	1.06
Sabit n=500	BIC	64.33	48.17	0.00	0.00	35.67	48.17	100	0.00	0.00	0.00
	CAIC	55	50.49	0.00	0.00	45	50.49	100	0.00	0.00	0.00
	AIC	95	2.53	4.83	2.48	0.16	0.41	100	0.00	10.46	2.56
	AIC3	92.66	16.51	0.33	0.52	7	16.67	100	0.00	1.72	1.31

^a Sadece eşdeğer olmayan modellere ilişkin doğru sonuçlardır.

^b Sadece eşdeğer modellere ilişkin yanlış pozitif sonuçlardır.

Hem eğim hem sabit parametresinin yanlış olduğu koşullara ait analiz bulguları incelendiğinde, örneklem büyüklüğünün bu koşullar için de etkili olduğu görülmektedir. Örneklem büyüklüğü 100'den 500'e yükseldiğinde tüm bilgi kriterleri için doğru karar oranı da yükselmiştir. Ölçek düzeyindeki analizlerde yanlış negatif karar oranı -özellikle örneklem büyüklüğü 100 olduğunda- yüksek olan bilgi kriterleri diğer tüm koşullarla benzer şekilde BIC ve CAIC olarak bulunmuştur. Bu bilgi kriterleri ile yapılan seçimlerdeki doğru karar oranı örneklem büyüklüğü yükseldiğinde artsa da AIC ve AIC3 ile alınan %100'lük doğru karar oranından düşük seviyededir. Madde düzeyindeki analizlerde benzer bulgulardan söz edilebilir. AIC ile alınan kararların yanlış pozitif olma oranı yine diğer koşullarla benzer şekildedir. Yanlılığın sadece sabit parametresinde olduğu koşullarda BIC ve CAIC ile alınan doğru karar oranı oldukça düşüktür. Hem eğim hem sabit parametrelerinin yanlış olduğu koşula oranla AIC ve AIC3 kullanıldığında da düşük doğru karar oranları mevcuttur. Yine de örneklem genişliği arttığında %90'nın üzerinde doğru karar oranları görülmektedir. Madde düzeyindeki analizler bu yanlışlık koşulu için de hem eğim hem sabit parametresinin yanlış olduğu koşul ile benzer şekildedir.

4. TARTIŞMA ve SONUÇ

Ölçme eşdeğerliğinin örtük sınıf modelleri ile incelenmesi oldukça yenidir ve alan yazında sınırlı sayıda örnek bulunmaktadır. ÖSA'nın diğer yöntemlerle ilişkisini inceledikleri çalışmada Kankaras, Vermunt ve Moors (2011) özellikle sabit parametresindeki yanlışlıkları tespit etmekte yanlışlık türüne de bağlı olarak çoklu-grup doğrulayıcı faktör analizinin yetersiz

kaldığını, madde tepki kuramı ve ÖSA'nın bu koşulda daha iyi sonuçlar verdiğini raporlamışlardır. Psikolojideki araştırmalarda kullanan süreksiz verilerin analizinde güçlü bir yaklaşım olarak öne çıktığından, ÖSA'nın psikolojik yapılara ilişkin ölçme eşdeğerli çalışmaları için değerli bir alternatif olduğu düşünülmektedir. Az sayıdaki simülasyon çalışmalarında da çalışmaların doğasından kaynaklı sınırlılıklar mevcuttur ve örtük sınıf modellerinde model seçim süreci ile ilgili de hala bir görüş birliği bulunmamaktadır. Örtük sınıf modellerinde bilgi kriterlerinin hassasiyetlerine ilişkin yapılan simülasyon çalışmalarının pek çoğu doğru sınıf sayısının belirlenmesine odaklansa da temelde yapılan işlem uygun modeli seçme olduğundan bu çalışma ile karşılaştırılabilir olduğu düşünülmektedir. Yang ve Yang (2007) simülasyon çalışmalarının bir parçası olarak 200, 400, 600, 800 ve 1000 örneklem büyüklüklerinin doğru sınıf sayısının belirlenmesi üzerine etkisini incelemişlerdir. Çalışmanın sonucunda BIC, CAIC ve AIC3 örneklem büyüklüğü arttıkça daha doğru sonuçlar verdiği, AIC'nin ise örneklem büyüklüğüne göre farklılaşmayan sonuçlar verdiği raporlanmıştır. BIC ve CAIC bilgi kriterleri için beklenen sonuç da yapıları itibarıyla (her iki bilgi kriteri hesaplanırken $\log(n)$ düzeltmesi kullanılmaktadır) bu yöndedir. Araştırmacılar örneklem büyüklüğü formülünde bulunmayan AIC3'ün de daha iyi sonuçlar vermesinin ilginç olduğuna değinmişlerdir. Bu çalışmada da benzer şekilde AIC3 bilgi kriteri özellikle ölçek düzeyindeki, analizlerde örneklem büyüklüğü arttığında %100'e yakın doğru sonuçlar verirken, 100 örneklem büyüklüğünde bu oran %67'dir. Lin (2012), kayıp veri oranı, toplam madde sayısı, örneklem büyüklüğü gibi farklı değişkenlerin etkisini incelediği simülasyon çalışmasında, 50, 100, 250 ve 500 örneklem büyüklüklerini kullanmıştır. Diğer çalışmalarla benzer olarak örneklem büyüklüğü arttıkça, doğru sonuç oranının da arttığı tespit edilmiştir. Sadece AIC bilgi kriterinde örneklem büyüklüğünün bir etkisi olmamıştır. Araştırmacı küçük örneklem büyüklüğü ve az madde sayısında özellikle CAIC'nin düşük doğru karar oranı olduğunu belirtmiştir. Bu çalışmada da dört bilgi kriteri içerisinde ölçek düzeyindeki analizlerde her iki örneklem büyüklüğünde de CAIC en az doğru karar veren bilgi kriteri olarak bulunmuştur. Ancak madde düzeyindeki analizlerde örneklem büyüklüğü 500'e yükseldiğinde CAIC de ortalama %99 gibi çok yüksek oranda doğru sonuçlar vermiştir. Bu çalışmanın örneklem büyüklüğüne ilişkin sonuçlarının benzer olduğu bir diğer simülasyon çalışması Kankaras, Vermunt ve Moors (2011)'un ölçme eşdeğerliği çalışmalarında farklı yöntemlerin etkililiğini inceledikleri çalışmadır. Yazarlar çalışmanın bir parçası olarak; AIC ve BIC bilgi kriterlerini, iki karşılaştırma grubu ve grup başına 200 ve 1000 örneklem büyüklüğünü ölçme eşdeğerliği kapsamında ele almışlar ve 200 örneklem büyüklüğünde hem ölçek hem madde düzeyindeki analizlerde AIC'nin daha iyi sonuçlar verdiğini, BIC'nin yanlışlığı saptamada yeterli olmadığını raporlamışlardır. Örneklem büyüklüğü 1000'e yükseldiğinde ölçek düzeyinde yine AIC daha doğru sonuçlar verirken; BIC'nin madde düzeyinde daha etkili sonuçlar verdiği tespit edilmiştir. Bu çalışmanın örneklem büyüklüğüne ilişkin bulguları da diğer çalışmalarla (Dias, 2006; Nylund, Asparouhov ve Muthen, 2007; Yang ve Yang, 2007) benzer şekilde örneklem büyüklüğü arttığında doğru modelin tercih edilme oranının arttığını göstermektedir. Örneklem büyüklüğünden en fazla etkilenen bilgi kriterleri ise sırasıyla BIC ve CAIC olarak tespit edilmiştir. AIC3 ve AIC için ise bu oran çok daha düşüktür. Bu nedenle küçük örneklemelerde AIC ve AIC3 bilgi kriterlerinin kullanılmasının daha uygun olduğu düşünülmektedir.

Çalışmanın eşdeğer olmayan madde sayısına ilişkin bulguları, Kankaras, Vermunt, ve Moors (2011)'un çalışmalarıyla paralellik göstermektedir. Toplam madde sayısının 5 olduğu koşulda; eşdeğer olmayan madde sayısı 1'den 3'e yükseldiğinde ölçek düzeyindeki analizlerde doğru karar oranı artış göstermektedir. Madde düzeyindeki analizlerde ise eşdeğer olmayan madde sayısının herhangi bir etkisi bulunmamaktadır. Bu çalışmada da özellikle Kankaras, Vermunt ve Moors (2011)'un çalışmasında olduğu gibi örneklem büyüklüğünün 500'e yükseldiği koşullarda (yazarların çalışmasında $n=1000$ şeklindedir) beş maddeden birinin yanlış olduğu koşul ile üçünün yanlış olduğu koşul arasında BIC bilgi kriterinin ölçek düzeyinde

oldukça farklı sonuçlar verdiği görülmektedir. Eşdeğer olmayan madde oranı arttığında BIC çok daha fazla doğru sonuç vermektedir. Bu çalışmada ek olarak toplam madde sayısının 10 olduğu koşullar da incelenmiştir. Ölçek düzeyindeki analizlerde, 10 maddeden birinin eşdeğer olmadığı koşulda doğru karar oranı, üç eşdeğer olmayan maddeye oranla oldukça düşük bulunmuştur. Yanlı madde sayısı arttığında ölçek düzeyinde tespit oranının da artması, en sade modelin aranması ve tahminlenen parametre sayısı düşünüldüğünde beklenen bir durumdur. Örneğin 10 maddeden birindeki sabit parametresinde yanlılık olduğunda, ölçek düzeyindeki analizlerde kısmi homojen model için tahminlenen parametre sayısı 91 iken bu sayı homojen modelde 51'e düşmektedir. Madde düzeyinde bir maddenin sabit parametresine kısıt konduğunda ise bu sayı 87 olmaktadır. Bu durumda Dias (2006)'ın çalışmasında belirttiği gibi BIC ve CAIC'nin daha sade modeli seçme eğilimi ile tutarlı sonuçlar elde edilmiş ve bilgi kriterleri daha sade olan yanlış negatif kararı işaret etmişlerdir. Benzer şekilde Yang ve Yang (2007) da CAIC'nin daha az parametre tahminlenen modeli seçme eğiliminde olduğunu belirtmişlerdir. Ancak beklenenin aksine örneklem büyüklüğü 500 olduğunda ölçek düzeyindeki 10 maddeden üçünün yanlı olması ile 5 maddeden üçünün yanlı olması koşulları arasında AIC ve AIC3 bilgi kriterleri kullanıldığında herhangi bir fark tespit edilmemiştir. BIC kullanıldığında da sadece %10'luk bir fark bulunmaktadır. CAIC bilgi kriteri ise eşdeğer olmayan madde oranı arttıkça-Yang ve Yang (2007)'ın çalışmalarına benzer şekilde- daha çok doğru sonuçlar vermiştir. Madde düzeyindeki analizlerde ise eşdeğer olmayan maddeler çok büyük oranda tespit edilebilmiştir. AIC ve AIC3 kullanıldığında eşdeğer maddelerin de yanlı olarak tespit edilme oranları mevcut olsa da bu oran oldukça düşüktür. Bu bulgu da diğer çalışmalara (Dias, 2006; Nylund, Asparouhov ve Muthen, 2007) benzer şekilde AIC'nin olması gerekenden fazla sınıf sayısı tespit etme eğilimine benzer bir nitelik taşımaktadır.

Dias (2006), 5 ve 8 madde kullandığı simülasyon çalışmasında madde sayısı arttığında AIC'nin olması gerekenden fazla sınıf sayısı olan modelin seçtiğini belirtmiştir. Bu çalışmada ise madde sayısının etkisi, maddelerin tamamının eşdeğer ya da yanlı olduğu yani homojen ve heterojen modele ait bulgular üzerinden ölçek düzeyinde değerlendirildiğinde, beş ve on madde olmasının sonuçlar üzerinde bir değişikliğe neden olmadığı görülmüştür. Benzer şekilde Nylund, Asparouhov ve Muthen (2007)'in çalışmalarında 8,10 ve 15 maddeden oluşan farklı madde sayılarının var olduğu koşullarda madde sayısının bir etkisi bulunmazken, koşullu olasılık örüntülerinin zorluk kolaylık derecelerinin doğru model seçimini farklılaştırdığı tespit edilmiştir. Bu çalışmanın yanlılığın olduğu parametrelere ilişkin bulguları incelendiğinde, ölçek düzeyindeki analizlerde hem eğim hem sabit parametresinin yanlı olduğu durumlarda yanlılığın tespit edilme oranının daha yüksek olduğu görülmektedir. Araştırmada kullanılan sabit ve eğim değerleri incelendiğinde, sabit değerlerindeki fark ile eğim değerlerindeki farkın olasılık dağılımında benzer sonuçlara yol açtığı görülmektedir. Bunun doğal sonucu olarak hem eğim hem sabit parametresi değişimlendiğinde tespiti daha kolay (olasılıklar arasında daha çok farkın olduğu) bir modelin varlığından söz edilebilir. Bu durumda yazarlarınkine (Nylund, Asparouhov ve Muthen, 2007) benzer şekilde ölçek düzeyindeki analizlerde daha kolay model daha yüksek oranda doğru şekilde tahminlenebilmiştir. Madde düzeyindeki analizlerde ise yanlılığın olduğu parametre değil fakat örneklem büyüklüğü doğru karar oranı üzerinde etkili görünmektedir. Kullanılan tüm bilgi kriterleri madde düzeyinde örneklem büyüklüğü 500 olduğunda %100'e yakın doğru sonuçlar vermektedir. Bu koşul için vurgulanması gereken nokta AIC ile yapılan seçimlerde tüm diğer koşullarla benzer şekilde yanlış pozitif karar olasılığının varlığıdır.

Bulguların geneli incelendiğinde bilgi kriterlerinden BIC ve CAIC'nin örneklem büyüklüğünden çok etkilendiğini ve yanlış negatif karar eğilimi olduğu söylenebilir. AIC ile alınan kararlar ise örneklem büyüklüğünden etkilenmemekte fakat yanlış pozitif karar eğilimi gözlenmektedir. Bu koşullarda en çok doğru karar alınan AIC3 olmasa da en tutarlı bilgi kriteri olarak öne çıkmaktadır. Bu bulgu Dias (2006)'ın çalışmasını desteklemektedir.

4.1. SONUÇ ve ÖNERİLER

Çalışmanın bulguları ışığında, ÖSA'nın süreksiz yapıdaki ölçme eşdeğerliği çalışmaları için uygun bir yöntem olduğu yorumu yapılabilir. Yöntemin özellikle farklı koşullarda oldukça yüksek ve doğru şekilde yanlılığı yakalama oranı dikkat çekicidir. Diğer yöntemlerle benzer şekilde örneklem büyüklüğünün artmasının model seçiminde olumlu katkısı olduğu söylenebilir. Araştırmacılar ölçek geliştirme, adaptasyon gibi çalışmalarda genellikle büyük örneklemle çalışmaktadır. Ancak küçük örneklemelerin varlığında geçerli grup karşılaştırmaları yapabilmek için yapılan ölçme eşdeğerliği çalışmalarında özellikle AIC doğru karar oranı en yüksek bilgi kriteri olarak öne çıkmaktadır. Fakat daha dikkatle incelendiğinde yanlış pozitif oranı en yüksek olan bilgi kriterinin de AIC olduğu fark edilmektedir. Bu noktada AIC3'ün tercih edilmesi uygun görülmektedir. Küçük örneklemelerde, BIC ve CAIC bilgi kriterleri yanlılığı tespit etmede yeterli olamamış ve yüksek oranda yanlış negatif sonuçlar vermiştir. Bu nedenle mümkün olduğunca geniş örneklemelerde çalışmak bir alternatif olsa da, küçük örneklemelerin varlığında BIC, CAIC kullanılmamasının daha geçerli sonuçlar için önemli olduğu düşünülmektedir.

Çalışmanın bir diğer önemli bulgusu ölçek düzeyinde ve madde düzeyinde yapılan analizlerin, özellikle eşdeğer olmayan madde oranı azaldığındaki farkıdır. Pratikte ÖSA ile yapılan görgül çalışmalarda araştırmacıların grup karşılaştırmaları yapmadan önce inceledikleri eşdeğerliğin ölçek düzeyinde olduğu bilinmektedir. Genellikle kısmi homojen model veriye uyum sağlamazsa, madde düzeyinde sabit parametrelerinin incelenmesi yolu tercih edilmektedir. ÖSA ile yapılan analizlerde BIC, CAIC ve AIC3 sadece bir maddede yanlılık olan koşullarda, bu yanlılığı ölçek düzeyinde tespit etme konusunda yeterli görülmemektedir. AIC ile yapılan ölçek düzeyindeki analizler daha doğru sonuçlar vermektedir. Fakat AIC kullanıldığında da yanlış pozitif karar olasılığı mevcuttur. Bu nedenle küçük örneklemeler söz konusu olduğunda ÖSA ile yapılan çalışmalarda AIC3'ün kullanılması önerilmektedir.

Yapılan çalışmalardan bazıları (örn. Yang ve Yang, 2007) BIC ve CAIC' de hesaplanırken kullanılan formüldeki $\log(N)$ kısmına yazılacak değerin N sayısına uygulanacak bir düzeltmeyle küçük örneklemelerde daha doğru sonuçlara ulaşılabileceğini belirtmektedirler. $(N+2)/24$ şeklinde olan bu düzeltme ile elde edilen BIC*, CAIC* şeklinde belirtilen bilgi kriterleri ile de çalışmalara ihtiyaç vardır.

5. KAYNAKLAR

- Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, 8, 338-363.
- Bornovalova, M. A., Levy, R., Gratz, K., & Lejuez, C. W. (2010). Understanding the Heterogeneity of BPD Symptoms Through Latent Class Analysis: Initial Results and Clinical Correlates Among Inner-City Substance Users. *Psychological Assessment*, 233-245.
- Brown, R. S. (2007). Using latent class analysis to set academic performance standards. *Educational Assessment*, 283-301.
- Campbell, S. B., Morgan-Lopez, A. A., Cox, M. J., & McLoyd, V. C. (2009). A latent class analysis of maternal depressive symptoms over 12 years and offspring adjustment in adolescence. *Journal of Abnormal Psychology*, 479-493.
- Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis with applications in the social, behavioral, and health sciences*. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Dias, J. G. (2006). Latent class analysis and model selection. M. R. Kruse, C. Borgelt, A. Nürberger, & W. Gaul içinde, *From data and information analysis to knowledge engineering* (pp. 95-102). Berlin: Springer-Verlag.
- Eid, M., Langeheine, R., & Diener, E. (2003). Comparing typological structures across cultures by multigroup latent class analysis : A primer. *Journal of Cross-Cultural Psychology*, 34, 195-210.
- Güngör, D., Korkmaz, M., Somer, O. (2013). Çoklu-grup örtük sınıf analizi ve ölçme eşdeğerliği, *Türk Psikoloji Dergisi*, 28 (72), 48-57.
- Güngör-Culha, D., & Korkmaz, M. (2011). Örtük sınıf analizi ile bir örnek uygulama. *Eğitimde ve Psikolojide Ölçme Değerlendirme Dergisi*, 2(2):191-199.

- Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18, 117-144.
- Kankaras, M., & Moors, G. (2009). Measurement equivalence in solidarity attitudes in Europe. Insights from a multiple group latent class factor approach. *International Sociology*, 24(4):557-579.
- Kankaras, M., Moors, G., & Vermunt, J. K. (2010). Testing for measurement invariance with latent class analysis. E. Davidov, P. Schmidt, & J. B. Billiet içinde, *Cross-Cultural Analysis: Methods and Applications*. Routledge.
- Kankaras, M., Vermunt, J. K., & Moors, G. (2011). Measurement equivalence of ordinal items. A comparison of factor analytic, item response theory, and latent class approaches. *Sociological Methods & Research*, Vol 40-2: 279-310.
- Klonsky, D. E., & Olino, T. M. (2008). Identifying clinically distinct subgroups of self-injurers among young adults: a latent class analysis. *Journal of Consulting and Clinical Psychology*, 22-27 doi:10.1037/0022-006X.76.1.22.
- Lin, T. H. (2006). A comparison of model selection indices for nested latent class models. *Monte Carlo Methods and Applications*, 12(3-4): 239-259.
- Lin, T. H. (2012). Model selection information criteria in latent class models with missing data and contingency question. *Journal of Statistical Computation and Simulation*, DOI:10.1080/00949655.2012.698621.
- Magidson, J., & Vermunt, J. K. (2004). Comparing latent class factor analysis with the traditional approach in data mining. H. Bozdogan içinde, *Statistical data mining and knowledge discovery* (pp. 373-383). Boca Raton: Chapman&Hall/CRC.
- McCutcheon, A. L. (2002). Basic concepts and procedures in single-and multiple-group latent class analysis. J. A. Hagenaars, & A. L. McCutcheon içinde, *Applied latent class analysis* (pp. 56-89). New York: Cambridge University Press.
- McCutcheon, A. L., & Hagenaars, J. A. (1997). Comparative social research with multi-sample latent class models. J. Rost, & R. Langeheine içinde, *Applications of latent trait and latent class models in the social sciences* (s. 266-277). Münster: Waxmann.
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York: Taylor & Francis Group.
- Moors, G., & Wennekers, C. (2003). Comparing moral values in western european countries between 1981 and 1999. A multigroup latent-class factor approach. *International Journal of Comparative Sociology*, 44:155-172 doi:10.1177/002071520304400203.
- Morren, M., Gelissen, J., & Vermunt, J. K. (2011). Dealing with extreme response style in cross-cultural research: a restricted latent class factor analysis. *Sociological Methodology*.
- Morren, M., Gelissen, J., & Vermunt, J. K. (2012). The impact of controlling for extreme responding on measurement equivalence in cross-cultural research. *Methodology*, 8(4), 159-170. doi:10.1027/1614-2241/a000048
- Nylund, K. L., Asparouhov, T., & Muthen, B. O. (2007). Deciding on the number of classes in latent class analysis in latent class analysis and growth mixture modeling: a monte carlo simulation study. *Structural Equation Modeling*, 14(4): 535-569.
- Somer, O., Korkmaz, M., Dural, S., & Can, S. (2009). Ölçme eşdeğerliğinin yapısal eşitlik modellemesi ve madde cevap kuramı kapsamında incelenmesi. *Türk Psikoloji Dergisi*, 24 (64):61-75.
- Silvia, P. J., Kaufman, J. C., & Pretz, J. E. (2009). Is creativity domain-specific? Latent class models of creative accomplishments and creative self-descriptions. *Psychology of Aesthetics, Creativity, and the Arts*, 139-148 doi:10.1037/a0014940.
- Vermunt, J. K., & Magidson, J. (2005). *Bilgisayar Yazılımı*. Statistical Innovations Inc. Belmont, MA.
- Vermunt, J. K., & Magidson, J. (2008). *LG-Syntax User's Guide: Manual for Latent GOLD 4.5 Syntax Module*. Belmont, MA: Statistical Innovations Inc.
- Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike information criterion and the Bayesian information criterion. *Psychological Methods*, 17:2, 228-243.
- Waasdorp, T. E., & Bradshaw, C. P. (2011). Examining student responses to frequent bullying: a latent class approach. *Journal of Educational Psychology*, 336-352, doi:10.1037/a0022747.
- Yang, C., & Yang, C. (2007). Separating latent classes by information criteria. *Journal of Classification*, 24, 183-203.

Extended Abstract

Measurement equivalence is described as having a true score independent from being a member of any group (Horn ve McArdle, 1992; Millsap, 2011). For example; if measurement equivalence is hold then males and females who have same ability should have same true scores. Multi-group confirmatory factor analysis in the context of structural equation modeling and differential item function in the context of item response theory are the most popular methods in detecting measurement (in)equivalence (Somer, Korkmaz, Dural & Can, 2009). Latent class analysis is a new method for investigating measurement equivalence.

Latent class analysis can be described as a latent variable model and it is usually used when both observed and latent variables are discrete. While investigating measurement equivalence with LCA three models are tested named as; heterogeneous, partial homogenous and homogenous models. In heterogeneous model all parameters- latent class probabilities and conditional probabilities- are free and estimated independently for all comparison groups (McCutcheon & Hagnaars, 1997). If heterogeneous model is found as the best model, it can be said that the difference is qualitative and groups cannot be compared. However, in partial homogenous models some parameters such as, some direct effects or interaction effects, are restricted. Especially, models in which all interaction effects are restricted to 0 is important. This model is similar to metric equivalence in structural equation models. In item response theory terms it is similar to models in which all slope parameters are equal for comparison groups that means relationship between observed and latent variables are equal for all groups (Kankaras, Moors & Vermunt, 2010; McCutcheon & Hagnaars, 1997). Another model, structural homogenous model is also named as homogenous model (Kankaras & Moors, 2009). This model is similar to scalar equivalence in structural equation models or in terms of item response theory models in which both slope and intercept parameters are equal. Practically if homogenous model is selected as the best model, group comparisons can be made.

Model selection is controversial in latent class analysis. There is a growing body of simulation studies about this procedure. An alternative is using likelihood ratio test (LRT) statistics; however there is a consensus about risks of using this method as it might not follow chi-squared distribution especially when the data is sparse (Collins and Lanza, 2010; Dias, 2006; Morren, Gelissen and Vermunt, 2011; Yang and Yang, 2007). An alternative is using information criteria like AIC, BIC. Some researchers suggest using AIC3 (for ex. Dias, 2006) while some others suggesting BIC (for ex. Nylund, Asparouhov & Muthen, 2007). The aim of current study is to investigate the performance of different information criteria while detecting measurement inequivalence with LCA.

28 different conditions, 100 replications per condition; totally 2800 data sets were simulated and analyzed. While generating data sets different homogeneity levels (heterogeneous, partial homogeneous, homogeneous) were taken into account due to investigate power of LCA detecting inequivalence in 3 different models. Hereby, just false positive or true positive decision rates and both these wrong decisions could be examined in the measurement equivalence studies with LCA. For an example of false negative decision; homogeneous or partial homogeneous model is selected although the heterogeneous model is more informative for the data, in other words inequivalence isn't determined while it should be. Similarly, false positive decision refers selecting partial homogeneous or heterogeneous model while homogeneous model should have been selected. When the true model is partial homogeneous there can be three different outcomes (true, false positive, false negative). Conditions in which both slope and intercepts are inequivalent represent heterogeneous models. Partial homogeneous models are characterized by inequivalent intercepts. Conditions in which both slope and intercepts are equivalent in other words inequivalent item numbers are zero represents homogeneous models. Different conditions were related to sample size (100 or 500), number of items in a scale (5 or 10), number of inequivalent numbers (0,1,3, 5 or 10), type of inequivalence (intercept, both slope and intercept, none). Following variables are fixed for all conditions: number of groups (2), sample size (100 or 500 per group), number of latent classes (2), latent class probabilities (.50), number of category numbers (5). There were 5 steps of our procedure. Data sets were simulated. Analyses were done both in scale and item levels. Information criteria values were calculated. Model selections were done. True, false negative and false positive decision rates were determined. Latent GOLD 4.5 were used for simulation and analyses.

Generally for all information criteria increasing in sample size induced increasing in true decision rate. While using BIC, CAIC it is recommended not to use small sample sizes. Number of items did not cause any alteration in true decision rates. In scale level analysis true decision rates increase parallel with the number of inequivalent items. Also inequivalence both in slope and intercept parameters had affirmative effect on choosing true models relatively inequivalence in only intercept parameter. In conclusion BIC and CAIC tend to select false negative, AIC tends to select false positive models hence AIC3 seems to be more consistent. This result is parallel with the results of Dias's (2006) study.

Ek.1. Simülasyon verisinin koşullu olasılık dağılımlarıEğim Parametresi Eşdeğer Olan ve Olmayan Maddelere Ait Koşullu Olasılıklar⁴

Eğim maddeler	Parametresi	Eşdeğer	Madde Kategorileri				
			1	2	3	4	5
Grup	Örtük Sınıf						
1	1		0.66	0.09	0.15	0.09	0.02
		2	0.01	0.01	0.08	0.34	0.57
		1	0.83	0.04	0.04	0.04	0.04
2	2		0.01	0.00	0.02	0.12	0.86
Eşdeğer olmayan maddeler							
Grup	Örtük Sınıf						
1	1		0.38	0.08	0.23	0.23	0.08
		2	0.03	0.02	0.15	0.40	0.40
		1	0.94	0.03	0.02	0.01	0.01
2	2		0.00	0.00	0.01	0.08	0.92

Eğim Parametresi Eşdeğer, Sabit Parametresi Eşdeğer Olan ve Olmayan Maddelere Ait Koşullu Olasılıklar

		Madde Kategorileri				
Eşdeğer maddeler		1	2	3	4	5
Grup	Örtük Sınıf					
1	1	0.66	0.09	0.15	0.09	0.02
	2	0.01	0.01	0.08	0.34	0.57
	1	0.66	0.09	0.15	0.09	0.02
2	2	0.01	0.01	0.08	0.34	0.57
Eşdeğer olmayan maddeler						
Grup	Örtük Sınıf					
1	1	0.66	0.09	0.15	0.09	0.02
	2	0.01	0.01	0.08	0.34	0.57
	1	0.83	0.04	0.04	0.04	0.04
2	2	0.01	0.00	0.02	0.12	0.86

⁴ Tablodaki koşullu olasılık toplamaları basamakların yuvarlanmasından kaynaklı 1'den farklı sonuçlar vermektedir.