



Ankara Üniversitesi Yabancı Uyruklu Öğrenci Seçme Testinin Ölçme Değişmezliğinin Örtük Sınıf ve Rasch Modeline Göre İncelenmesi *

Özge Altıntaş¹, Ömer Kutlu²

Öz

Bu araştırmanın genel amacı, Ankara Üniversitesi Yabancı Uyruklu Öğrenci Sınavı'nda yer alan Temel Öğrenme Becerileri Testi'nin faktör yapılarını belirleyerek testin ülkelere ve cinsiyete göre ölçme değişmezliğine sahip olup olmadığını ve testte yer alan maddelerin madde işlev farklılığı gösterip göstermediğini ortaya koymaktır. Bu araştırma betimsel araştırma modellerinden tarama türündedir. Araştırmanın genel evrenini 2134 birey oluşturmaktadır. Araştırma, kültürler arası karşılaştırmaları içerdiği için kültür, örneklem seçiminde ölçüt alınmıştır. Buna göre, Almanya, Azerbaycan ve İran'dan seçilen 1110 birey örnekleme oluşturmuştur. Araştırmada Temel Öğrenme Becerileri Testi'nin ilk 60 maddelik harf, sayı ve şekil ilişkilerine dayalı bölümü kullanılmıştır. Testin faktör yapısını belirlemede Örtük Sınıf Analizi ve Örtük Sınıf Faktör Analizi, ülkelere ve cinsiyete göre ölçme değişmezliğine sahip olup olmadığını ortaya koymada Eşanlı Örtük Sınıf Analizi kullanılmıştır. Testte yer alan maddelerin ülkelere ve cinsiyete göre madde işlev farklılığı gösterip göstermediğini belirlemede ise, Rasch Modeline dayalı Özyinelemeli Bölümlenme Analizi kullanılmıştır. Temel Öğrenme Becerileri Testi'nin faktör yapısını belirlemeden önce testin bir bütün olarak ölçek yapısı eğilimi gösterip göstermediği, testi alan bireylerin yanıt örüntüleri açısından hangi sınıflarda yer aldıkları incelenmiştir. Buna göre, testin dört sınıflı örtük bir modele sahip olduğu ve birden fazla faktör yapısı içerdiği görülmüştür. Bu faktörler sırasıyla, parça-bütün ilişkisi kurma, çıkarım yapma ve analitik düşünme olarak adlandırılmıştır. Testin ülkelere ve cinsiyete göre ölçme değişmezliğine ilişkin bulguları incelendiğinde ise, testin homojen bir yapıya sahip olduğu bulgusuna ulaşılmıştır. Sonuç olarak testin ölçtüğü yapı, karşılaştırılan gruplar açısından eşdeğerdir. Testte yer alan maddelerin ülkelere ve cinsiyete göre madde işlev farklılığı gösterip göstermediğine bakıldığında ise, ülkeler açısından 14 maddede madde işlev farklılığına rastlanmıştır. Cinsiyet açısından ise madde işlev farklılığına rastlanmamıştır.

Anahtar Kelimeler

Test geliştirme
Örtük sınıf faktör analizi
Ölçme değişmezliği
Eşanlı örtük sınıf analizi
Madde işlev farklılığı
Rasch modeline dayalı
özyinelemeli bölümlenme analizi
Uluslararası öğrenci seçme testleri

Makale Hakkında

Gönderim Tarihi: 19.04.2019
Kabul Tarihi: 20.11.2019
Elektronik Yayın Tarihi: 04.05.2020

DOI: 10.15390/EB.2020.8685

* Bu makale Özge Altıntaş'ın Ömer Kutlu danışmanlığında yürüttüğü "Ankara Üniversitesi Yabancı Uyruklu Öğrenci Seçme Testi'nin Ölçme Değişmezliğinin Örtük Sınıf ve Rasch Modeline Göre İncelenmesi" başlıklı doktora tezinden üretilmiştir.

¹ Ankara Üniversitesi, Eğitim Bilimleri Fakültesi, Eğitim Bilimleri Bölümü, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Türkiye, oaltintas@ankara.edu.tr

² Ankara Üniversitesi, Eğitim Bilimleri Fakültesi, Eğitim Bilimleri Bölümü, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Türkiye, omerkutlu@ankara.edu.tr

Giriş

Bir toplumda siyasal, ekonomik ve kültürel sistemin gelişerek devamlılığını sağlayacak bireyler eğitim kurumları tarafından yetiştirilmektedir. Eğitimin toplumsal gelişmeyi sağlayan bu işlevi tüm toplumlarda eğitim kurumlarının önemini arttırmıştır. Toplumlar bireylerin kazanmaları gereken temel yaşam becerilerini özellikle okulöncesi, ilkokul ve ortaokul süreçlerinde sağlarken; liseyi ağırlıklı olarak temel zihinsel becerilerin gelişimi ve mesleğe yönelme süreci olarak görmektedirler. Yükseköğretim programları ise, bir yandan meslek seçme ve o meslekte yetkinleşmeyi sağlarken bir yandan da bireylerin yükseköğrenim düzeyinde entelektüel, aktüel ve bilimsel bilgi birikiminin de zenginleşmesini sağlamaktadır.

Son yıllarda, liseden mezun olan birçok öğrenci, daha kaliteli bir üniversite eğitimi almak amacıyla kendisini ülkesiyle sınırlandırmamakta, ülkesi dışında bulunan üniversitelerin çeşitli yükseköğretim programlarında okumak istemektedir. Türkiye’de, yükseköğretim programlarında öğrenim görmek isteyen yabancı uyruklu öğrencilerin seçilmesi ve yerleştirilmesi işlemleri; 1980 yılı ve öncesi, 1981-2011 yılları arası, 2011 yılı ve sonrası olmak üzere üç grupta ele alınarak incelenebilir. 1980 yılı ve öncesinde üniversiteler, yabancı uyruklu öğrenci seçme ve yerleştirme işlemlerini kendilerinin belirlediği ölçütlere göre yapmaktaydı. 1981-2011 yılları arasında yabancı uyruklu öğrenci seçme ve yerleştirme işlemlerinde ise Yabancı Uyruklu Öğrenci Sınavı (YÖS) testleri kullanılmaktaydı. Ancak Yükseköğretim Kurulu (YÖK) 2010 yılında aldığı karar uyarınca yabancı uyruklu öğrencilerin seçilmesi ve yerleştirilmesi işlemlerini üniversitelerin kendilerine bırakmıştır. Bu karar uyarınca üniversiteler, açacakları programları ve kontenjan sayılarını YÖK’ten onay alarak oluşturmuşlardır (YÖK, 2013). Ankara Üniversitesi Rektörlüğü de, Ankara Üniversitesi Ölçme ve Değerlendirme Uygulama ve Araştırma Merkezi (ANKÜDEM) bünyesinde, Ankara Üniversitesi Yabancı Uyruklu Öğrenci Sınavı’nı (AYÖS) 2011 yılından itibaren yapma kararını almıştır (ANKÜDEM, 2011).

Psikolojik özellikleri ölçmek amacıyla geliştirilmiş testlerin ele aldığı zihinsel yapılar ve bu yapıların hangi boyutlardan oluştuğu bir yapı geçerliği sorunudur. Bu durum test geliştirme süreci açısından önemlidir. Test geliştiriciler, testi öncelikle tanımladıkları mantıksal yapıya uygun olarak geliştirirler; daha sonra bu yapının varlığını istatistiksel olarak doğrulamaya çalışırlar. Testin yapı geçerliğini belirlemeyi amaçlayan bu süreç test geliştirmede, testin faktör yapılarının belirlenmesi olarak adlandırılmaktadır (Bollen, 1989). Psikolojik yapıları ölçen testler üzerinde yapılan çalışmalar, testin bütününden daha çok, oluşturduğu alt boyutlar ve maddeler düzeyinde yürütülmektedir. Bu çalışmalar testten elde edilen puanların güvenilir ve geçerli sonuçlar ortaya koymasına amacıyla yapılmaktadır. Bu nedenle AYÖS gibi uluslararası öğrencilerin katılımıyla gerçekleştirilen sınavlarda kullanılan testlerin bazı psikometrik özelliklerinin incelenmesi gerekmektedir. AYÖS testinin ölçtüğü psikolojik yapılar ve bu testlere katılan öğrenci özellikleri dikkate alındığında, bir bütün olarak testin öğrenci özelliklerine göre değişip değişmediğinin araştırılması bir gereklilik olarak ortaya çıkmaktadır.

Psikolojik yapıları ölçen testler, bu yapıların altında yatan örtük özellikleri ölçmeye çalışmaktadır. Testlerin ölçmeyi amaçladığı özellikleri grup özelliklerinden etkilenmeden ölçebilmesi önemlidir ve bu durum alanyazında ölçme değişmezliği (measurement invariance) kavramıyla ele alınıp incelenmektedir. Testi oluşturan maddelerin ölçmeyi amaçladığı özellik dışındaki etkenlerden etkilenerek amacı dışında farklı roller üstlenmesi ise alanyazında madde işlev farklılığı (differential item functioning) kavramıyla incelenmektedir (Mellenbergh, 1989; Meredith, 1993; Millsap, 2011; Shealy ve Stout, 1993; Vandenberg ve Lance, 2000). Ölçme değişmezliği ve madde işlev farklılığı (MIF) kavramları, psikolojik özellikleri ölçen testlerin uluslararası düzeyde yaygın kullanımıyla birlikte psikometrinin ve ölçme biliminin konuları arasında yanlılık çalışmaları içinde önemli bir yer tutmaya başlamıştır.

Amerikan Eğitim Araştırmaları Derneği (American Educational Research Association [AERA]), Amerikan Psikoloji Derneği (American Psychological Association [APA]) ve Eğitimde Ölçme Ulusal Konseyi (National Council on Measurement in Education [NCME]), testlerin ve test uygulamalarının kalitesini ölçmek ve test etiği oluşturmak amacıyla 1966 yılında, bugünkü bilinen adıyla eğitsel ve psikolojik ölçmeler için bazı standartlar oluşturmuşlardır. 1974, 1985, 1999 ve son olarak 2014 yıllarında revizyonu yapılan *Eğitimde ve Psikolojide Ölçme Standartları* raporunda teknik bir terim olarak kullanılan test yansızlığı (test fairness/fairness as lack of bias), *ölçülen yapı açısından eşit düzeyde olan bireylerin içinde buldukları gruptan bağımsız olarak bir testten aynı puanları elde etmesi* olarak ifade edilmektedir (AERA, APA ve NCME, 2014).

Ankara Üniversitesi bünyesinde hazırlanan AYÖS testleri adı geçen rehberde yer alan birçok standardı ölçüt olarak hem test geliştirmede hem de farklı alt gruplardan gelen, farklı dil ve kültür özelliklerine sahip bireyler için yansız bir test uygulaması gerçekleştirmede uluslararası düzeyi yakalamayı amaçlamıştır. Çünkü uluslararası düzeyde, farklı demografik ve kültürel özelliklere sahip bireylere uygulanan testlerden elde edilen puanların demografik ya da kültürel özelliklerden bağımsız, yansız ve adil olması testlerin geçerliği açısından bir gerekliliktir. Bireylerin testlerden almış olduğu puanlar arasında karşılaştırmalı çalışmalar yapabilmek bu gerekliliğin yerine getirilmiş olmasına bağlıdır. Dünyanın farklı ülkelerinden bireylerin katılımıyla gerçekleştirilen AYÖS'te kullanılan test ile bu testi oluşturan maddelerin, demografik ve kültürel özellikler açısından farklılık gösterip göstermediğinin belirlenebilmesi ölçme değişmezliği ve madde işlev farklılığı çalışmalarının yapılmasını gerektirmektedir.

Buna göre bu araştırmanın genel amacı, AYÖS'te yer alan Temel Öğrenme Becerileri Testi'nin (TÖBT) faktör yapılarını belirleyerek testin ülkelere ve cinsiyete göre ölçme değişmezliğine sahip olup olmadığını ve testte yer alan maddelerin madde işlev farklılığı gösterip göstermediğini ortaya koymaktır.

Bu amaç doğrultusunda aşağıdaki sorulara yanıt aranmıştır:

1. TÖBT'ün Almanya'ya, Azerbaycan'a ve İran'a göre faktör yapısı nasıldır?
2. TÖBT,
 - a) ülkelere (Almanya, Azerbaycan ve İran) ve
 - b) cinsiyete göre ölçme değişmezliğine sahip midir?
3. TÖBT'ü oluşturan maddeler,
 - a) ülkelere (Almanya, Azerbaycan ve İran) ve
 - b) cinsiyete göre madde işlev farklılığı göstermekte midir?

Psikolojik ölçme araçları, ister dile dayalı ister dilden bağımsız geliştirilmiş olsunlar, bireylerin yaş ve cinsiyet gibi demografik; etnik köken ve ülke gibi kültürel özelliklerinden etkilenmektedir (Messick, 1989). Bu durum ölçme araçlarından elde edilen puanlara dayalı olarak verilen kararların doğruluğunu sorgulanır duruma getirmektedir. Bu nedenle, bir ölçme aracının ölçmek istediği psikolojik yapıyı bu özelliklerden bağımsız olarak ölçtüğünden emin olmak gerekir. Bunun için bir testin, testi yanıtlayan bireylerin ait olduğu gruplar açısından yapısal denkliği ve gruplardan kaynaklanabilecek farklılıklar incelenmelidir. Bu çalışma AYÖS'te kullanılan TÖBT'ün testi alan bireylerin dahil oldukları gruplar açısından yapısal denkliğini ve bu gruplardan kaynaklanabilecek farklılıkları incelemesi açısından önemlidir.

Araştırmanın önemli görülen bir başka yönü ise, bu çalışmanın 1980'li yıllardan beri Türkiye'de uygulanan yabancı uyruklu öğrenci seçme ve yerleştirme sınavlarında kullanılan testlere ilişkin ilk çalışma olmasıdır. Türkiye'nin yükseköğretimdeki eğitim standartlarını artırma çabası (Bologna süreci, Avrupa Birliği değişim programları, akreditasyon çalışmaları vb.), uluslararası öğrencilerin dikkatini çekmektedir. Uluslararası öğrencilerin seçiminde kullanılan testlerin güvenilir ve geçerli kararlar

verebilmesi, yükseköğretim süreçlerine akademik açıdan başarıları yüksek olan öğrencilerin katılmasını sağlayacaktır. Bu çalışmanın, Türkiye’de uluslararası düzeyde uygulanan bir sınava ilişkin ilk çalışma olmasının yanı sıra bu türden sınavlarda kullanılan testlerin ve testlerde yer alan maddelerin eşdeğerliğine yönelik ilk çalışma olması açısından önem taşımaktadır. Bunun yanında, çalışmadan elde edilen bulgular, hem test geliştirme süreçleri hem de ölçme sonuçlarını tehdit eden durumların ortadan kalkmasını sağlaması açısından önemlidir.

Yöntem

Araştırma Modeli

Bu araştırma, bir testin psikometrik özelliklerini olduğu biçimiyle betimlemeyi amaçlamadığından betimsel araştırma modellerinden tarama türündedir. Bu çalışmada betimleme, testte yer alan maddelere ilişkin bilgilerin anlamını güçlendirmek, dikkat çekmek ve bilgilerin özüne dokunmadan alıntılarının yapılmasını sağlamak amacı taşımaktadır (Wolcott, 1994). Bu çalışmanın da konusu olan testin psikometrik özellikleri kendi koşulları içinde olduğu gibi betimlemeye çalışılmıştır.

Evren ve Örneklem

Araştırmanın evrenini 2013 yılında AYÖS’e giren 944’ü kadın 1190’ı erkek olmak üzere toplam 2134 birey oluşturmaktadır. Bu araştırma genel evreni oluşturan 2134 bireyden seçilmiş örneklem üzerinde yürütülmüştür. Çalışmanın konusuna bağlı olarak, derinlemesine araştırma yapmak ve araştırmanın amacı hakkında zengin bilgilere ulaşmak için örneklem seçiminde amaçlı örnekleme yöntemlerinden ölçüt örnekleme kullanılmıştır. Bununla birlikte bu araştırma kültürlerarası karşılaştırmaları içerdiği için kültür (ülke), örneklem seçiminde ölçüt olarak dikkate alınmıştır. Bu nedenle hem öğrenci sayılarının yüksekliği hem de farklı kültürleri temsil ettiği için Almanya, Azerbaycan ve İran örneklem olarak belirlenmiştir. Örneklemi oluşturan grubun ülkelere göre dağılımı Tablo 1’de verilmiştir.

Tablo 1. Örneklemi Oluşturan Grubun Ükelere ve Cinsiyete Göre Dağılımı

Ülke	Kadın		Erkek		Toplam	
	Frekans (f)	Yüzdeler (%)	Frekans (f)	Yüzdeler (%)	Frekans (f)	Yüzdeler (%)
Almanya	129	66.15	66	33.85	195	17.57
Azerbaycan	157	28.49	394	71.51	551	49.64
İran	172	47.25	192	52.75	364	32.79
Toplam	458	41.26	652	58.74	1110	100.00

Tablo 1’e bakıldığında, örnekleme yer alan 1110 bireyden 458 (yaklaşık %41)’i kadın; 652 (yaklaşık %59)’u erkektir. Ayrıca örnekleme, Azerbaycan’dan 551 (yaklaşık %50); İran’dan 364 (yaklaşık %33) ve Almanya’dan 195 (yaklaşık %17) birey yer almaktadır. Guilford (1954), örtük değişken modellerine ilişkin çözümlenmelerde örneklem sayısının 200 olmasının yeterli olacağını belirtmektedir. Buna göre bu çalışmada kullanılan çözümlenmeler için örneklem sayısının yeterli düzeyde olduğu görülmektedir.

Veriler ve Toplanması

Veriler, Türkiye (Ankara), Almanya (Köln) ve Azerbaycan (Bakü) olmak üzere üç farklı sınav merkezinde, tek oturumda ve eş zamanlı olarak uygulanan AYÖS TÖBT’e verilen yanıtlardan oluşmaktadır. Veriler düzenlenirken iki farklı kitapçık türüne göre elde edilen yanıtlar, A kitapçık türü temel alınarak 1-0 puan matrisine çevrilip birleştirilerek çözümlenmeye hazır duruma getirilmiştir.

Veri Toplama Aracı

Araştırma kapsamında kullanılan veri toplama aracı, kendi içinde iki alt bölümden ve 100 maddeden oluşan bir akademik yetenek testidir. Testin ilk 60 maddelik harf, sayı ve şekil ilişkilerine dayalı bölümü genel olarak, analitik düşünme, akıl yürütme, soyut ve uzaysal düşünme gibi zihinsel

özellikleri ölçen maddelerden oluşmaktadır. Diğer 40 madde ise, matematik ve geometri bilgilerinin kullanımını gerektiren sayısal düşünme becerilerini ölçen maddelerden oluşmaktadır (ANKÜDEM, 2012). Bu araştırmada, ölçtüğü özellikler bakımından daha benzeşik olması ve daha sınırlı sayıda maddeyi içermesi açısından TÖBT'ün ilk 60 maddelik kısmı ele alınmıştır.

Testte yer alan maddelerin yazımında testin ölçtüğü özellikler dikkate alınmakta ve maddeler bu özelliklere göre testi geliştiren uzmanlar tarafından gruplandırılmaktadır. Maddelerin belirtilen özelliklere dağılımına, hazırlanan belirtke tablosuyla karar verilmektedir. Buna göre belirtke tablosu; bütüne uygun olan/olmayan parçayı bulma, bir bütün içinde verilen şekil, sayı ve benzeri bilgilerden sonuç çıkarma, verilen ilişkilerden, şekillerden yararlanarak ilişkilerin dayandığı kuralı ya da sayıyı bulma gibi özellikler ve bu özellikleri ölçen maddelerden oluşmaktadır. Test, son formunda yer alacak madde sayısının en az üç katı kadar madde yazılarak geliştirilmektedir. Testin son formu, belirtke tablosu dikkate alınarak ölçülmeye çalışılan psikolojik özelliği en iyi ölçtüğü düşünülen maddeler seçilerek oluşturulmaktadır. Bu yolla akılcı ve mantıksal süreçlerle testin hem kapsam geçerliği hem de yapı geçerliği kontrol altına alınmaya çalışılmaktadır.

TÖBT'ten elde edilen puanlara ilişkin güvenilirlik değerlerinin bulunduğu betimsel istatistikler hem evren hem de örneklem sayısına göre Tablo 2'de verilmiştir.

Tablo 2. Evren ve Örneklem Grubun Betimsel İstatistikleri ($k = 60$)

Grup	N	Aritmetik Ortalama	Standart Sapma	KR-20 Güvenirliği
Evren	2134	45.05	9.47	0.91
Örneklem	1110	45.48	9.01	0.90

Tablo 2'ye bakıldığında, testi alan tüm gruba ve bu araştırma için çözümlenmeler yapıldığı örneklem grubuna ait aritmetik ortalama ve standart sapma değerlerinin birbirine yakın olduğu görülmektedir. Bu durum, örneklemin evreni temsil ettiğini göstermektedir. Bunun yanı sıra, teste ilişkin hem evrenden hem de örneklemden elde edilen KR-20 güvenilirlik değerlerinin birbirine yakın ve yüksek olduğu görülmektedir. Bu güvenilirlik değerinin tercih edilme nedeni testte yer alan maddelerin güçlük indekslerinin birbirlerinden farklı olmasıdır (Kuder ve Richardson, 1937).

Verilerin Çözümlemesi

Araştırmada ilk olarak, Almanya, Azerbaycan ve İran'ın oluşturduğu örneklem açısından TÖBT'ün faktör yapısı Örtük Sınıf Faktör Analizi (ÖSFA)'yla belirlenmiştir. TÖBT'ün ülkelere ve cinsiyete göre ölçme değişmezliğine sahip olup olmadığı, Eşanlı Örtük Sınıf Analizi (EÖSA) kullanılarak ortaya koyulmuştur. Bu analizlerin kullanılma nedeni, verilerin elde edildiği testin yapısının sıralama düzeyinde olması ve testte yer alan maddelerin ikili (binary) yanıtlara sahip olmasıdır. Çözümlenmeler için Latent GOLD 5.1 programının sırasıyla, Basic ve Advanced/Syntax sürümlerinden yararlanılmıştır (Vermunt ve Magidson, 2013a, 2013b, 2013c).

Testte yer alan maddelerin ülkelere ve cinsiyete göre madde işlev farklılığı gösterip göstermediğinin belirlenmesinde ise, Rasch modeline dayalı olarak geliştirilen Özyinelemeli Bölümleme Analizi (ÖBA) kullanılmıştır. Çözümlenmeler, açık kaynak kodlu bir istatistik programı olan R programında yer alan psychotree paketinin 0.12-1 sürümü kullanılarak yapılmıştır (Zeileis, Strobl, Wickelmaier ve Kopf, 2011).

Aşağıda araştırmada kullanılan çözümlenmelere ilişkin kısa bir bilgi verilmiştir.

Örtük Sınıf Faktör Analizi: ÖSFA, ortak değişkenlik kaynağını paylaşan değişkenlerin bir araya gelmesiyle oluşan iki ya da daha fazla kategoriye sahip faktörleri tanımlamaktadır. Alanyazında geleneksel Örtük Sınıf Analizi (ÖSA) ile geleneksel Faktör Analizi (FA)'yı birleştiren modeller örtük sınıf faktör modelleri altında, ÖSFA olarak yer almaktadır (Magidson ve Vermunt, 2001, 2003a, 2003b, 2004; Vermunt ve Magidson, 2000, 2005a).

ÖSA ile ÖSFA arasındaki amaçsal benzerlikler Bartholomew, Steele, Moustaki ve Galbraith (2008) tarafından şu şekilde açıklanmaktadır: (1) Gözlenen değişkenler arasındaki ilişkileri açıklamak ve keşfetmek, (2) Bu ilişkilerin daha az sayıda örtük değişkenle açıklanıp açıklanamayacağını belirlemek ve (3) Yanıtlar temelinde her bir örtük değişken için her bir bireyin puanını saptamak.

ÖSFA, daha az sayıda değişken kullanarak FA'dan daha iyi performans gösterebilmektedir. FA'da, en az üç sürekli düzeyde değişkene gerek vardır; bu üç değişken ise yalnızca tek bir faktörü tanımlayabilmektedir. ÖSFA'da, üç dikotom (iki kategorili) değişken benzer şekilde bir faktör üretebilmektedir. Ancak, örtük sınıf modelleri yalnızca dikotom değişkenlerle sınırlı değildir; ortak değişkenlerin (covariates) de modele dahil edilmesiyle ek faktörler tanımlanabilmektedir (Magidson ve Vermunt, 2003a).

Amaçsal benzerliklerinin ve kullandıkları veri matrisinin yapısı açısından farklılaşmalarının yanında FA'nın ÖSFA'ya göre uygulamada dört önemli sınırlılığı bulunmaktadır. Bunlar; (1) tüm değişkenlerin sürekli olması, (2) çok değişkenli normallik ve doğrusallık gibi varsayımlara sahip olması, (3) örtük değişkenlerin (faktörlerin) eşit aralıklı ya da oranlı ölçek düzeyinde olması ve (4) çözümlerin yorumlanmasında araştırmacıya olası rotasyon seçeneklerinden birini seçmek zorunda bırakması, yani tek bir sonucunun bulunmamasıdır (Magidson ve Vermunt, 2003b). Bu yaklaşım sayesinde FA'nın yukarıda sözü edilen dört sorunu, yöntemin kendi sınırlılıkları dahilinde çözülebilmektedir.

Örtük Sınıf Faktör Analizi'nde parametre kestirimi diğer örtük sınıf modellerinde olduğu gibi Beklentinin Maksimizasyonu (Expectation Maximization [EM]) ya da Newton-Raphson (NR) algoritmasını kullanan En Çok Olabilirlik (Maximum Likelihood [MLH]) yöntemiyle ya da bu iki algoritmanın bileşimini kullanan MLH yöntemiyle yapılmaktadır. Dolayısıyla ÖSFA'nın tek dezavantajı, kestirilen parametrelerin yorumlanması aşamasında ortaya çıkmaktadır. Çünkü ÖSFA, FA'da yer alan faktör yük değerleri, faktör-madde korelasyonları, faktör korelasyonları ve ortak varyanslar gibi kolay yorumlanabilir parametrelere sahip değildir. Bu sorun için Vermunt ve Magidson (2005a), ÖSFA modeliyle birlikte elde edilen MLH kestirimlerinin doğrusal bir yaklaşımla kullanılmasını önermektedirler. Bu yaklaşımın kullanılması çıktılarının FA'dakine benzer şekilde elde edilmesini mümkün kılmakta; bu sayede daha güvenilir ve doğrusal olmayan bir faktör analitik modeli kullanılarak ölçülmek istenen özelliğin altında yatan örtük faktör yapıları belirlenebilmektedir.

Eşanlı Örtük Sınıf Analizi: Gruplar arasındaki örtük yapıların karşılaştırılabilmesi için Clogg ve Goodman (1984, 1985), geleneksel örtük sınıf analizinin çok gruplu bir uzantısı olan Çoklu Grup Örtük Sınıf Analizi (ÇGÖSA) yaklaşımını önermişlerdir. Bu yöntem sayesinde, iki ya da daha fazla gruba ilişkin gözlenen değişkenden yola çıkarak örtük değişkenlerin örtük yapıları çözümlenebilmektedir. Bu çözümlenelerde birden fazla grup, sahip oldukları örtük özellikler açısından karşılaştırıldığı gibi tek bir grubun farklı zaman dilimlerindeki örtük özellikleri de karşılaştırılabilmektedir. Bu sayede gruplar arasında ölçme değişmezliği incelemeleri de yapılabilmektedir.

Alanyazında çoğunlukla, EÖSA ya da ÇGÖSA olarak da ifade edilen bu yöntemde yer alan modeller, olasılıksal, log-lineer ya da lojistik parametrelendirmeler olmak üzere üç farklı parametrelendirme kullanılarak gerçekleştirilmektedir (Kankaras ve Vermunt, 2014). Hem gösterge değişkenleri hem de örtük değişken/ler sınıflama düzeyinde olduğunda yalnızca olasılıksal parametrelendirme kullanılabilirdiğinden bu araştırmada eşanlı örtük sınıf modeli, olasılık parametreleriyle açıklanmıştır. Buna göre gruplar arasında ölçme değişmezliği, belirli bir sınıfta yer alan madde yanıt olasılıklarının eşitlenmesiyle incelenmektedir (Clogg ve Goodman, 1985).

Eşanlı Örtük Sınıf Analizi araştırmacılara, grup farklılıkları olduğu durumda, farklılığın nerede olduğunu ve bu farklılıkların nasıl yorumlanacağı hakkında bilgi vermektedir (Collins ve Lanza, 2010). Dolayısıyla araştırmacıların gruplar arası karşılaştırmalarda, örtük sınıf sayısının eşit, koşullu

olasılıkların eşdeğer ve örtük sınıf yapılarının benzer olup-olmadığını ortaya koyması gerekmektedir. Belirtilen durumlar açısından farklılıklar belirlendiğinde ise, bu farklılıkların nasıl açıklanabileceğinin belirtilmesi gerekmektedir.

Ölçme değişmezliği incelemeleri EÖSA'yla farklı homojenlik düzeylerinde yapılmaktadır. Çünkü gruplar arasında örtük yapıların karşılaştırılmasında bir dizi olası sonuç ortaya çıkmaktadır. Buna göre gruplar, birbirlerinden tamamen ayrışik (heterojen model), kısmen benzeşik (kısmi homojen) ya da birbirleriyle tamamen benzeşik (homojen model) olabilmektedirler (Kankaras, Moors ve Vermunt, 2011). Model seçiminde ise, diğer örtük sınıf analizlerinde olduğu gibi kay-kare test istatistikleri ve bilgi kriterleri kullanılmaktadır. Diğer ölçme değişmezliği inceleme yöntemlerinde olduğu gibi eşanlı örtük sınıf analiziyle yapılan ölçme değişmezliği incelemelerinde de amaç, değişmezlik düzeyi en yüksek olan modeli seçmektir.

Ölçme değişmezliğinin EÖSA'yla incelenmesinde, değişmezlik incelemelerinin çeşitli düzeylerde eşdeğerliğe sahip modellerin karşılaştırılmasına dayandığından model karşılaştırmalarında yapılan parametre kestirimleri de tıpkı geleneksel örtük sınıf analizlerinde olduğu gibi, EM ve NR algoritmalarının kullanıldığı MLH yöntemiyle yapılmaktadır. Model karşılaştırmalarında model uyumunun değerlendirilmesinde ise yine, Akaike Bilgi Kriteri (AIC) ve Bayesian Bilgi Kriteri (BIC) gibi bilgi kriterlerinin yanı sıra kay-kare istatistiklerinden yararlanılmaktadır.

Kankaras, Moors ve Vermunt (2011), kültürler arası çalışmalarda ölçme değişmezliğinin test edilmesi için örtük sınıf analizinin çeşitli avantajlara sahip olduğunu belirterek bu anlamda oldukça değerli bir yaklaşım sunduğunu vurgulamaktadırlar. Araştırmacılara göre bu durumun üç önemli nedeni bulunmaktadır: Bunlardan ilki, araştırmalarda kullanılan ölçme araçlarında yer alan maddelerin çoğunlukla ayrık (sıralama ya da sınıflama) yanıt kategorilerine sahip olması ve örtük sınıf analizinin bu türden ayrık değişkenler arasındaki ilişkilerden yola çıkarak örtük yapıları tanımlamada kullanılabilir olmasıdır. İkincisi, sıklıkla kullanılan Çoklu Grup Doğrulayıcı Faktör Analizinden (ÇGDFA) farklı olarak örtük sınıf modelinde, örtük değişkenlere sınıflama düzeyindeymiş gibi yaklaşıyor olmasıdır. Örneğin, bir grup kategorik değişkenden yola çıkarak tipolojik bir sınıflandırma yapılabilmesi gibi. Benzer şekilde bir veri setinde yer alan kategorik değişkenlerin ölçeklenebilirliğini araştırmak amacıyla değişkenlere sıralama düzeyindeymiş gibi de yaklaşabilmektedir. Üçüncü önemli neden ise, örtük sınıf analizine dayalı yöntemlerin, bu yöntemlere göre daha yaygın olarak kullanılan ve daha güçlü dağılım varsayımları gerektiren ÇGDFA ile çok gruplu MTK yaklaşımlarına göre daha esnek bir alternatif öneriyor olmasıdır.

Kültürler arası karşılaştırmalı çalışmalarda EÖSA'nın kullanılmasının avantajlarından biri, analizin geleneksel yaklaşımlara göre daha esnek oluşudur. Örneğin, ÇGDFA'da gruplar arası karşılaştırma yapabilmek için, her bir faktör altında ölçme değişmezliğine sahip en az iki maddenin yer alması gerekirken; EÖSA'da ölçme değişmezliğine sahip olmayan madde sayısı fazla olsa bile gruplar arası karşılaştırmalar yapılabilir. Bunun yanı sıra, ÇGDFA'da elde edilen bulguları ölçme değişmezliği açısından yorumlayabilmek için en az kısmi değişmezlik düzeyinin sağlanmış olması gerekirken; EÖSA'da yapısal değişmezlik sağlanmasa bile gruplar arası karşılaştırmalara ilişkin yorumlama yapabilmek mümkündür (Kankaras ve Moors, 2009). Ancak bu noktada araştırmacıların çoğunlukla tam ölçme değişmezliğiyle ilgilendikleri ve özellikle de grupların karşılaştırılabilirliğini tam olarak sağlayabilmeyi istediklerini belirtmekte yarar vardır. Dolayısıyla örtük sınıf analizi bağlamında bunu yapabilmek için yapısal eşdeğerliğin mutlaka sağlanması gerekmektedir (Hagenaars ve McCutcheon, 2002).

Rasch Modeline Dayalı Özyinelemeli Bölümleme Analizi: Zeileis, Hothorn ve Hornik (2008), MİF'in Rasch modeli çerçevesinde belirlenmesinde, yorumlamayı güçleştirmeden hem önceden belirlenen gruplara hem de olası tüm gruplara ilişkin testleri içeren Model Tabanlı Özyinelemeli Bölümleme (Model Based Recursive Partitioning [MBRP]) adı verilen yeni bir istatistiksel yaklaşım

önermişlerdir. Bu sayede kurulan modeldeki parametre dengesizlikleri saptanabilmektedir. Örtük sınıf ya da karışım modellerine benzer biçimde, modelde yer alan parametrelerin farklılaştığı grupları tanımlamaya dayalı bu yaklaşımın altında yatan ana düşünce, MİF'e neden olabilecek olası tüm kaynakları araştırarak tüm grupların ardışık bir biçimde test edilmesidir. Özyinelemeli bölümlenme gruplar, örtük sınıf modellerinde olduğu gibi örtük bir faktör tarafından değil, gözlenen ortak değişkenlerin kombinasyonlarıyla, sezgisel yaklaşım¹ kullanılarak tanımlanmaktadır. Böylece, modele dayalı bölümlenme, örtük sınıflara ya da karışım modellerine alternatif olarak sezgisel, ancak yorumlanması kolay alternatifler sunabilmektedir.

Gruplar arası karşılaştırmalı çalışmalarda, MİF'i belirlemede oldukça yeni olan bu yöntem, ekonometri alanından uyarlanan yapısal değişim testlerinin kullanıldığı model tabanlı özyinelemeli bölümlenme yöntemine dayanmaktadır. Model tabanlı özyinelemeli bölümlenme, kategorik ya da sürekli düzeydeki yanıt değişkeninin farklı değerlere sahip olduğu grubu tanımlamak için ortak değişken alanının yinelemeli olarak bölümlendirildiği sınıflandırma ve regresyon ağaçları yöntemleriyle oldukça ilişkilidir. Yarı parametrik bir yaklaşıma sahip olan bu yöntemde, tek bir yanıt değişkenine ilişkin değerler yerine gruplar arasında değişen parametrik bir modelin parametreleri yer almaktadır. Bu tür parametreler, gruplar arasında değişen Rasch modelinin madde parametreleri olabileceği gibi doğrusal regresyon modelinin sabit ve eğim parametreleri de olabilmektedir (Strobl, Kopf ve Zeileis, 2015).

Model tabanlı özyinelemeli bölümlenme yöntemiyle yapılan çözümlenmelerde amaç, veri matrisini kendi içinde homojen bir yapı gösteren alt gruplara (sınıflara) ayırmaktır. Bu alt grupların her birine node (boğum/düğüm) adı verilmektedir. Alt gruplar öncelikle, ortak değişkenler (yaş ve cinsiyet gibi) aracılığıyla tanımlanır. Daha sonra bu boğumlar, kendi içinde özdeş bir yapıya sahip oluncaya kadar, sınıflandırma ve regresyon ağaçlarında olduğu gibi bir ağacın dalları gibi ayrılmaya başlar. Bu ağaç Rasch ağacı (Rasch tree) olarak tanımlanmakta ve bu ağacın dallarında ortak değişkenlere ilişkin kritik değerler (yapraklar) yer almaktadır. Süreç, her bir boğum içi varyans en düşük değeri alınca kadar; boğumlar arası varyans en yüksek değeri alınca kadar devam etmektedir (Kopf, Augustin ve Strobl, 2010).

Bu çerçevede her bir yaprağı, uygun bir modelle (en çok olabilirlik modeli ya da doğrusal regresyon modeli gibi) ilişkilendirilmiş boğumlara sahip bir Rasch ağacı oluşturularak model uyumu sağlanmaya çalışılır. Buradaki temel fikir, her bir boğumun tek bir modelle ilişkilendirilmiş olmasıdır. Aşağıda Rasch ağacı oluşturmak için kullanılan ardışık adımlar (algoritma) verilmiştir (Strobl, Wickelmaier ve Zeileis, 2011; Zeileis vd., 2008):

1. Örnekleme yer alan tüm bireyler için madde parametreleri ortak kestirilir.
2. Her bir ortak değişkene göre madde parametrelerinin kararlılığı değerlendirilir.
3. Madde parametreleri anlamlı düzeyde kararlı bir yapı göstermiyorsa (anlamlı düzeyde kararlılık sağlanmıyorsa) örneklem, model uyumu iyileşene kadar, ortak değişkenlere göre kesme noktasından bölünür.
4. Anlamlı düzeyde kararlılık sağlanana kadar (ya da alt örneklem çok küçük olduğunda) elde edilen alt örneklemde 1-3 arası adımlar yinelemeli olarak tekrar edilir. Bir başka anlatımla durdurma ölçütüne ulaşılan kadar bölümlenmeye devam edilir.

¹ Deneysel bir ispat yolu gerektirmeyen ancak karmaşık bir problemi keşfedici ve deneyimlere dayalı bir yolla daha basit bir hale getirerek en uygun çözümün elde edilmeye çalışıldığı yaklaşım.

Bulgular

Testin faktör yapısını ÖSFA yardımıyla ortaya koymadan önce, testin faktör yapısı hakkında ipucu sağlayacak bazı belirlemeler yapılmıştır. Bu amaçla, bireylerin teste verdikleri yanıtlara dayalı olarak oluşturulan sınıfların bir ölçek yapısı eğilimi gösterip göstermediği ÖSA'yla priori (önsel) olarak belirlenmiştir. Buna göre beklenen durum oluşturulan sınıflara göre, TÖBT'te yer alan bir maddenin herhangi bir kategorisindeki değerinin tutarlı biçimde artması; diğer kategorisindeki değerinin ise tutarlı biçimde azalmasıdır. Bu durum, testin bir ölçek yapısı oluşturabileceğine ilişkin işaret olarak yorumlanmaktadır (Vermunt ve Magidson, 2005b). Bunun için her biri farklı sayıda sınıfı belirten dört² farklı örtük sınıf modeli kestirilmiştir. Kestirim değerleri Tablo 3'te verilmiştir.

Tablo 3. TÖBT'e İlişkin Örtük Sınıf Model Kestirimi

Model	LL	BIC (LL)	AIC (LL)	Par. Sayısı	L ²	sd	p-değeri	Sın. Hatası
1-Sınıf	-31077.55	62575.82	62275.09	60	46662.88	1050	0.00	0.00
2-Sınıf	-28251.32	57351.10	56744.64	121	41010.43	989	0.00	0.02
3-Sınıf	-26876.07	55028.35	54116.15	182	38259.94	928	0.00	0.01
4-Sınıf	-26141.19	53986.33	52768.39	243	36790.18	867	0.00	0.02

Tablo 3'te, sınıf sayısının 1'den 4'e kadar olduğu modellerle bu modellere ilişkin uygunluk ölçüleri, parametre sayısı ve serbestlik derecesine ilişkin değerler yer almaktadır. Model seçim sürecine geçmeden önce kurulan tüm modellere ilişkin serbestlik derecesinin pozitif değere sahip olup olmadığına bakılması gerekmektedir. Çünkü negatif değere sahip serbestlik derecesi, örtük sınıf modelinin kestirim sürecinde tanımlanabilirlik sorununa yol açmaktadır (McCutcheon, 1987). Tablo 3'te bu değerlerin pozitif olduğu; bu durumda modelde yer alan parametre sayısının sütun rankına (sırasına) eşit olduğu belirtilebilir. Model değerlendirme sürecine uygunluk ölçülerinin yorumlanmasıyla devam etmek gerekirse bu ölçülerden L² istatistiği, model kestirimi yapıldıktan sonra açıklanamayan değişkenler arasındaki ilişkinin miktarını belirtmektedir. Değer ne kadar düşükse modelin veriye daha iyi uyum sağladığına ilişkin bir yorum yapılmalıdır. Tablodaki p değeri, L² istatistiğinin kay-kare dağılımına uygun olduğu varsayımıyla her bir model için manidarlık değerini gösterdiğinden örtük sınıf sayısının belirlenmesinde bu değere bakılmalıdır (Vermunt ve Magidson, 2005b). L² istatistiğinin kullanıldığı model seçim sürecinde model karşılaştırmaları en karmaşık iki model ile model kabulüne ilişkin hipotezler kurularak başlar. Hipotez testi sonucunda manidar bir farklılık bulunması durumunda, ilgili model uyumunun bir diğerine göre daha iyi olduğu belirtilir. Bu süreç tüm modeller test edilene ya da modeller arasında manidar bir farklılık bulunana kadar devam etmektedir (Lin, 2006). Model kabulüne ilişkin hipotezler ise aşağıdaki gibi kurulmaktadır:

H₀: Model uygundur.

H₁: Model uygun değildir.

Kurulan tüm modellerin manidarlık düzeyleri, belirlenen manidarlık düzeyi sınırının altında kaldığından ($p < 0.05$) H_0 hipotezi reddedilmiştir. Bu durum, kurulan dört sınıflı modelin uygun olmadığını göstermektedir. Ancak L² istatistiği, örneklem büyüklüğünden etkilenen bir değer olduğundan üzerinde çalışılan örneklem kay-kare dağılımına uygun olmadığında anlamlı sonuçlar üretmeyecektir. Bu nedenle, model uyumunun değerlendirilmesinde yalnızca L² istatistiğini kullanmak bu istatistiğin kay-kare dağılımı gerektiren ve örnekleme bağımlı yapısı nedeniyle çok da uygun değildir (McCutcheon, 2002). Dolayısıyla model uyumu değerlendirmelerini bilgi kriterleri gibi diğer uygunluk ölçüleriyle birlikte yapmak daha yerinde olacaktır.

Model seçiminde kullanılan bilgi kriterleri için de tıpkı L² istatistiğinde olduğu gibi elde edilen değer ne kadar düşükse, modelin veriye daha iyi uyum sağladığı yorumunun yapılması gerekmektedir. Buna göre Tablo 3'teki log-olabilirlik değerine bağlı hesaplanan BIC ve AIC değerleri incelenmiş; sınıf sayısı arttıkça değerlerin de giderek azaldığı görülmüştür. Ancak burada dikkat edilmesi gereken önemli bir nokta, değerlerin giderek azalıyor olmasının daha fazla sayıda sınıf içeren modelin seçilmesi

² Örtük sınıf analizinde genel bir başlangıç kuralı olarak sınıflar, 1 ile 4 arasında kestirilmektir (Vermunt ve Magidson, 2005b).

gerektiği anlamına gelmiyor olmasıdır. Şöyle ki, model kestirimine daha fazla sınıf içeren modellerle devam edildiğinde daha küçük bilgi kriteri değerlerine ulaşılsa da modelde yer alan parametre sayısı artacağından modeli yorumlamak zorlaşacaktır. Dolayısıyla bu kadar çok parametreye sahip bir model tutucu olma özelliğini kaybedecektir. Böylesi bir durumda Lin (2006) modelin kolay yorumlanabilmesi için daha az parametre içeren modelin tercih edilmesi gerektiğini belirtmektedir. Bu nedenle, TÖBT'e ilişkin model uyumu değerlendirmelerinde parametre sayısı da göz önünde bulundurulmuştur. Buna göre elde edilen dört sınıflı modelin hem BIC ve AIC değerleri hem de parametre sayısı açısından uygun olduğu söylenebilir.

Model seçimi yapıldıktan sonra üzerinde çalışılan örneklemin kay-kare dağılımına uymaması nedeniyle gerçek manidarlık değerinden uzaklaşan L^2 istatistiğinin manidarlık değeri bootstrapping yaklaşımıyla yeniden test edilebilmektedir. Bu sayede model seçim sürecinde, çok boyutlu bir karar verme yaklaşımı izlenmiş olmaktadır. L^2 istatistiğinin manidarlık değerinin bootstrap yöntemiyle yeniden elde edilebilmesi için öncelikle uygun modelin seçilmiş olması gerekmektedir. Çünkü bu yöntemde, seçilen modele ilişkin manidarlık değeriyle bu örneklemden hareketle rassal olarak üretilen örneklemden elde edilen manidarlık değeri karşılaştırılmaktadır. Bu işlem Latent GOLD programında yer alan *Bootstrap L^2* komutu sayesinde kolaylıkla yapılabilmektedir. Yeniden elde edilen manidarlık değeri ise, bir önceki değerden farklı olarak belirlenen manidarlık düzeyini aşmalı, yani 0.05 'ten büyük olmalıdır. Buna göre dört sınıflı model için bootstrapping yaklaşımıyla yeniden elde edilen L^2 istatistiğinin manidarlık değeri, 0.99^3 olarak bulunmuştur. Bu durumda $p > 0.05$ olduğundan kabul edilen dört sınıflı modelin daha sade ve dolayısıyla uygun olduğu söylenebilir. Tablo 4'te, TÖBT'ten elde edilen dört sınıflı modele ilişkin bazı sınıflandırma istatistikleri verilmiştir.

Tablo 4. Dört Sınıflı Modele İlişkin Sınıflandırma İstatistikleri

Sınıflandırma Hatası	0.02
Hata İndirgeme Değeri (λ)	0.97
Entropi R^2 Değeri	0.97
Standart R^2 Değeri	0.96

Tablo 4'te dört sınıflı modele ilişkin sınıflandırma hatasının oldukça düşük (%2) olduğu görülmektedir. Bu durumda modelin, %98 doğrulukla sınıflandırma yaptığı söylenebilir. Hata indirgeme, entropi R^2 ve standart R^2 değerleri ise gözlenen değişkenlerin örtük sınıflara ne kadar doğru atandığına ilişkin yapılan kestirimleri göstermektedir. Bu değerlerin 1.00 değerine yakın olması yapılan kestirimlerin iyiliğine işaret etmektedir. Tablo 4'te bu değerler sırasıyla, 0.97, 0.97 ve 0.96'dır. Buna göre yapılan kestirimlerin oldukça başarılı olduğunu söylemek mümkündür.

Teste ilişkin elde edilen dört sınıflı modeli açıklamada, örtük sınıf modelinin iki önemli parametresi olan örtük sınıf olasılıkları ile koşullu olasılıklardan yararlanılmıştır. Bu değerler aynı zamanda, TÖBT'ün faktör yapısına ilişkin bazı ipuçları içermelerinden dolayı önemlidir. Buna göre örtük sınıf sayısına ve bu sınıfların göreceli büyüklüklerine ilişkin bilgi veren örtük sınıf olasılıkları hesaplanmıştır. Toplamları 1.00'e eşit olan bu değerler sırasıyla, 0.31, 0.30, 0.21 ve 0.18 olarak elde edilmiştir. Yani örnekleme yer alan tüm bireylerin TÖBT'e ilişkin elde edilen dört sınıfa dağılımı Sınıf 1 için (π_1^X), 0.31; Sınıf 2 için (π_2^X), 0.30; Sınıf 3 için (π_3^X), 0.21 ve Sınıf 4 için (π_4^X), 0.18'dir. Buradan hareketle, bireylerin sınıflara dengeli bir biçimde atandıkları ifade edilebilir. Örtük sınıf olasılıklarının göreceli büyüklükleri farklılaştığı zaman, sınıfların dağılımı da ölçülen özellik açısından farklılaşmaktadır. Bu araştırma kapsamında yer alan bireylerin sınıflara dengeli biçimde dağılması ölçülen özellik açısından keskin bir farklılaşmanın olmadığına işaret etmektedir.

Sınıflardan herhangi birinde yer alan bireylerin TÖBT'te yer alan bir maddeye doğru ya da yanlış yanıt verme olasılığı koşullu olasılıklarla ifade edilmektedir. Koşullu olasılıklar, sınıfları birbirinden ayıran yanıt örüntüleri arasındaki farklılıkları göstermekte; bu yönüyle örtük değişkenin doğası hakkında bilgi vermektedir. Bu olasılık değerlerinin her bir maddenin yanıtlanma düzeylerine

³ Yöntem her seferinde rassal bir yolla örneklem üreteceğinden elde edilen bu değer bir sonraki seferde değişecektir. Ancak yorum, her kestirim için hemen hemen aynı olmaya devam edecektir (Vermunt ve Magidson, 2005b).

göre toplamı 1.00'e eşittir. Bu durum diğer sınıflarda yer alan bireylerin koşullu olasılık değerleri için de geçerlidir. Buna göre, TÖBT'teki maddelere verdikleri doğru ve yanlış yanıtlar açısından bireylerin profilini gösteren dört sınıf ve bu sınıflarda yer alan bireylerin bazı maddelere verdikleri yanıtlara ilişkin koşullu olasılık değerleri, aşağıda verilmiştir:

Sınıf 1, TÖBT'te yer alan maddeleri doğru yanıtlama olasılıkları yüksek bireyleri ifade etmektedir. Bu sınıf için beklenen durum, 0 düzeyine (yanlış yanıt) ilişkin koşullu olasılık değerlerinin düşük; 1 düzeyine (doğru yanıt) ilişkin koşullu olasılık değerlerinin yüksek olmasıdır. Örneğin Sınıf 1'de yer alan bir bireyin 14. maddeye doğru yanıt verme olasılığı (π_{211}^{14X}) 1.00 iken yanlış yanıt verme olasılığı (π_{111}^{14X}) 0.00'dır; 20, 32 ve 35. maddelere doğru yanıt verme olasılığı ($\pi_{211}^{20X} = \pi_{211}^{32X} = \pi_{211}^{35X}$) 0.99 iken yanlış yanıt verme olasılığı ($\pi_{111}^{20X} = \pi_{111}^{32X} = \pi_{111}^{35X}$) 0.01'dir.

Sınıf 2'de yer alan bireyler, TÖBT'te yer alan maddeleri yanıtlama olasılıkları bakımından Sınıf 1'e benzemektedir. Ancak, özellikle 46-50 arası maddelere Sınıf 1'de yer alan herhangi bir bireyden farklılaşarak 1 düzeyinde daha düşük olasılıkla doğru yanıt vermişlerdir. Örneğin bu sınıfta yer alan bir birey 12. madde için 0.99 (π_{212}^{12X}), 28. madde için 0.90 (π_{212}^{28X}) ve 41. madde için 0.93 (π_{212}^{41X}) gibi yüksek koşullu olasılık değerlerine sahipken; 46, 47, 48, 49 ve 50. maddeler için 0.03 ile 0.16 arasında düşük koşullu olasılık değerlerine sahiptir.

Sınıf 3'te yer alan bireyler, TÖBT'teki maddelere verdikleri doğru yanıtlar açısından Sınıf 1 ve 2'den farklılaşmışlardır. Bu sınıf 17-21 arası maddelere, diğer sınıflarda yer alan bireylerin yanıt örüntülerinden belirgin biçimde oldukça düşük olasılıkla doğru yanıt vermişlerdir. Bu maddeler 0.01 ile 0.03 arasında düşük koşullu olasılık değerlerine sahiptir. Ancak bu sınıfın tüm maddelere doğru yanıt verme olasılıkları daha düşük olsa da diğer sınıflara benzeyen bir yanıt verme eğilimine sahip olduğu görülmektedir.

Sınıf 4'te yer alan bireylerin TÖBT'teki tüm maddelere verdiği doğru yanıtların dağılımı diğer sınıflara benzemekle birlikte, bu sınıf diğer sınıflara göre oldukça düşük olasılıkla maddelere doğru yanıt vermişlerdir. Örneğin 5. madde 0.35; 25. madde 0.31; 30. madde 0.15 ve 46. madde 0.12 koşullu olasılık değerine sahiptir.

Koşullu olasılık değerleri bir bütün olarak incelendiğinde, maddelere verilen doğru yanıtlar açısından Sınıf 1, 2, 3 ve 4'te yer alan bireyler genel olarak benzer bir yanıt örüntüsüne sahipken özellikle 17-21 ve 46-50 arası maddelere verdikleri doğru yanıtlara ilişkin koşullu olasılık değerleri açısından bu dört sınıfta yer alan bireyler önemli ölçüde farklılaşmaktadırlar. Bu maddeler dışında kalan maddelere verdikleri doğru yanıtlar açısından bakıldığında ise genel olarak Sınıf 1, 2 ve 3'te yer alan bireyler göreceli olarak birbirine daha yakınken; Sınıf 4 diğer üç sınıfa göre farklılaşmaktadır.

Örtük sınıf analizi açısından belirlenmesi gereken bir diğer durum, maddelere doğru ya da yanlış yanıt veren birey yüzdelerinin her bir sınıf için hesaplanmasıdır. Sınıflandırma yüzdelerinin belirlenmesi, TÖBT'ün faktör yapısının ÖSFA'yla ortaya konulmasından önce testin faktör yapısı hakkında ipucu verecek bir işlemdir. Bu değerler, maddelere doğru ya da yanlış yanıt veren birey yüzdelerinin sınıflara dağılımı hakkında bilgi vermektedir. Bu yüzdeler sütun yüzdelerinin incelendiği koşullu olasılıklardan farklı olarak satır yüzdeleri açısından incelenmektedir. Bu nedenle, her bir madde düzeyinde yer alan oranların bulunduğu satırların toplamı 1.00'e eşittir. Yüzdelerin sınıflara dağılımında beklenen durum, yüzdenin herhangi bir madde düzeyindeki değerinin tutarlı biçimde artması; diğer düzeyindeki değerinin ise tutarlı biçimde azalmasıdır (Vermunt ve Magidson, 2005b).

Bu araştırmada da elde edilen sınıflar açısından beklenen durum, yüzde değerlerin herhangi bir maddenin 1 düzeyi (doğru yanıt) için, Sınıf 1'den 4'e doğru giderek azalması; 0 düzeyi (yanlış yanıt) için, Sınıf 1'den 4'e doğru giderek artmasıdır. Bu, psikolojik yapıları ölçmeyi amaçlayan bir testin tek boyutlu olabilmesi için istenen ideal bir durumdur. Maddelerin yanıt düzeylerine göre, sınıfların yüzde dağılımları sözü edilen durumu içermekle birlikte bazı farklılıklar göstermektedir. TÖBT'ün faktör yapısını belirleyebilmek amacıyla ÖSA'yla yapılan önsel belirlemelere bir bütün olarak bakıldığında teste yanıt veren bireylerin farklı yanıt örüntülerine sahip olduğu görülmüştür. Bu durum testin faktör yapısının birden fazla boyuta sahip olabileceğini işaret etmektedir.

TÖBT'ün Faktör Yapısı: TÖBT'ün faktör yapısını belirleyebilmek için ÖSFA yardımıyla farklı faktör sayılarını içeren modeller kurulmuştur. Başlangıç modeli olarak, ÖSA yardımıyla elde edilen dört sınıflı modelin bulgularından yararlanılmış, bir zorunluluk olmamasına rağmen testin faktör yapısının birden fazla boyuta sahip olabileceğine ilişkin önsel belirleme dikkate alınmıştır. Buna göre model seçim süreci, iki faktörlü başlangıç modeli kurularak başlatılmıştır. TÖBT'e ilişkin örtük sınıf faktör model kestirimlerine ilişkin değerler Tablo 5'te verilmiştir.

Tablo 5. TÖBT'e İlişkin Örtük Sınıf Faktör Model Kestirimi

Model	LL	BIC (LL)	AIC (LL)	Par. Sayısı	L ²	sd	p-değeri	Sın. Hatası
2-Faktör	-26381.65	54039.51	53127.30	182	37271.09	928	0.00	0.00
3-Faktör	-25488.28	52680.50	51462.55	243	35484.34	867	0.00	0.00
4-Faktör	-25950.83	53033.34	51509.66	304	36790.18	806	0.00	0.00

Tablo 5'te faktör sayısının 4'e kadar olduğu modeller ile bu modellere ilişkin uygunluk ölçüleri, parametre sayısı ve serbestlik derecesine ilişkin elde edilen değerler verilmiştir. Buna göre modellere ilişkin serbestlik derecesi değerlerinin pozitif olduğu görülmektedir. Bu durum modelde yer alan parametre sayısının sütun sırasına eşit, dolayısıyla tüm modellerin tanımlanabilir olduğunu göstermektedir.

Model kestirimlerine ilişkin Tablo 5'te yer alan uygunluk ölçülerine bakıldığında, hem bilgi kriterlerinin hem de L² değerinin iki faktörlü başlangıç modelinden sonra kurulan üç faktörlü modelde azaldığı görülmektedir. Ancak kurulan tüm modellere ait L² istatistiğinin manidarlık değeri, belirlenen manidarlık düzeyi sınırının altında kaldığından ($p < 0.05$) bu istatistik değeri için H_0 hipotezi reddedilmiştir. Bu durum, kurulan modelin uygun olmadığını göstermektedir. Ancak daha önce de belirtildiği gibi, L² istatistiği, örneklem büyüklüğünden etkilenmekte ve üzerinde çalışılan örneklem, kay-kare dağılımına uygun olmadığında anlamlı sonuçlar üretmemektedir. Bu nedenle, model uyumu değerlendirmelerini bilgi kriterleri gibi diğer uygunluk ölçüleriyle birlikte yapmak daha yerinde olacaktır. Buna göre Tablo 5'te yer alan BIC ve AIC değerleri incelenmiş ve faktör sayısı arttıkça bu değerlerin giderek azaldığı dört faktörlü modelde ise bir miktar yükseldiği görülmektedir. Bu durumda TÖBT'e ilişkin elde edilen üç faktörlü modelin hem BIC hem de AIC değeri açısından uygun olduğu söylenebilir. Tablo 6'da TÖBT'ten elde edilen 3-faktör modele ilişkin bazı sınıflandırma istatistikleri verilmiştir.

Tablo 6. 3-Faktör Modele İlişkin Sınıflandırma İstatistikleri

	Faktör 1	Faktör 2	Faktör 3
Sınıflandırma Hatası	0.00	0.04	0.03
Hata İndirgeme Değeri (λ)	0.99	0.92	0.87
Entropi R ² Değeri	0.99	0.86	0.86
Standart R ² Değeri	0.99	0.89	0.88

Tablo 6'da yer alan üç faktörlü modele ilişkin sınıflandırma istatistiklerinden sınıflandırma hatasının tüm faktörler için oldukça düşük (sırasıyla %0, %4 ve %3) olduğu görülmektedir. Bu durumda modelin, Faktör 1 için %100, Faktör 2 için %96 ve Faktör 3 için %97 doğrulukla sınıflandırma yaptığı söylenebilir. Hata indirgeme değeri, Faktör 1 için 0.99, Faktör 2 için 0.92 ve Faktör 3 için 0.87'dir. Entropi R² değeri, Faktör 1 için 0.99, Faktör 2 ve Faktör 3 için 0.86'dır. Standart R² değeri ise, Faktör 1 için 0.99, Faktör 2 için 0.89 ve Faktör 3 için 0.88'dir. Elde edilen bu değerlerin 1.00'e yakın ve oldukça yüksek olması yapılan kestirimlerin oldukça başarılı olduğunu göstermektedir.

Testin üç faktörlü yapısına ilişkin maddelere ait yük değerleri ise Faktör 1 için, 0.15 ile 0.94; Faktör 2 için, 0.08 ile 0.72; Faktör 3 için, 0.14 ile 0.53 aralığındadır. Maddelere ait yük değerleri, ilişkili olduğu maddenin R²'sine ya da ortak varyansına karşılık gelen kare kök değerini gösteren korelasyon katsayıları olarak rapor edilmektedir. Bu sayılar, doğrusal regresyon katsayıları olarak yorumlanabileceği gibi FA'daki faktör yük değerleri gibi de yorumlanabilmektedir.

Maddelere ilişkin R^2 değerleri ise, 0.01 ile 0.89 aralığındadır. Model tarafından açıklanan R^2 değeri, her bir maddenin testin bütününe ne düzeyde katkı yaptığını göstermektedir. Bu değer sıralı, sürekli ve sayılabilen değişkenler için standart R^2 ölçümlerini ifade etmektedir. Bu araştırmada olduğu gibi sınıflama düzeyindeki değişkenler için R^2 yorumu farklılaşmaktadır. Buna göre her bir kategori için dikotom yanıt değişkenleri gibi davranılan her bir R^2 , ölçümün ağırlıklandırılmış ortalamasını temsil eden Goodman-Kruskal τ -b katsayısı olarak ifade edilmektedir. R^2 , varyans analizindeki açıklanan varyansa, faktör analizindeki madde ortak varyanslarına benzetilmektedir (Vermunt ve Magidson, 2005b). Buna göre elde edilen değerler, örtük sınıf modellerinin genel çözümleme mantığı açısından her bir maddenin içinde yer aldığı faktöre ve genel olarak testin bütününe yani ölçülmek istenen psikolojik yapıya katkısı açısından ele alınarak yorumlanmıştır. Elde edilen faktörlerin adlandırılması için, maddelerin kendi içinde ölçtüğü özellikler açısından hem sayı şekil ve harf ilişkilerine dayalı madde yazan uzmanların hem de ölçme ve değerlendirme bilim alanındaki akademisyenlerin görüşlerine başvurulmuştur. Buna göre maddelerin ölçtüğü psikolojik özelliklerin faktörlere göre dağılımı Tablo 7’de verilmiştir.

Tablo 7. Maddelerin Ölçtüğü Psikolojik Özelliklerin Faktörlere Göre Dağılımı

Faktör Adı	Psikolojik Özellik	Madde Numarası	Toplam
Parça-Bütün İlişkisi Kurma	Bütüne uygun olan/olmayan parçayı bulma	1, 2, 17, 18, 19, 20, 21, 22, 23, 24, 25, 31	12
Çıkarım Yapma	Bir bütün içinde verilen şekil, sayı ve benzeri bilgilerden sonuç çıkarma	4, 5, 6, 9, 11, 26, 29, 32, 33, 46, 47, 48, 49, 50	14
Analitik Düşünme	Verilen ilişkilerden, şekillerden yararlanarak ilişkilerin dayandığı kuralı ya da sayıyı bulma	3, 7, 8, 10, 12, 13, 14, 15, 16, 27, 28, 30, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60	34
Toplam			60

Tablo 7’ye göre TÖBT, parça-bütün ilişkisi kurma, çıkarım yapma ve analitik düşünme boyutlarını ölçmeyi amaçlamaktadır. Maddelerin faktörlere dağılımına bakıldığında, Faktör 1 (12 madde) ve Faktör 2’nin (14 madde) birbirine yakın sayıda, Faktör 3’ün (34 madde) ise diğer iki faktöre göre daha fazla sayıda madde içerdiği görülmektedir. Bir test tek boyutlu değilse, ortaya çıkan her bir alt boyutun kendi içerisinde tanımlanmış özelliği ölçmesi beklenir. Bu nedenle alt boyutlar/faktörler arasındaki korelasyon değerlerinin negatif yönlü ya da düşük düzeyde olması beklenir. Bu araştırmada elde edilen faktörler arasındaki korelasyon değerleri şu şekildedir; Faktör 1 ile Faktör 2 arasında, -0.23; Faktör 1 ile Faktör 3 arasında, 0.25; Faktör 2 ile Faktör 3 arasında, -0.30’dur. Değerlerin tümü beklendiği üzere düşük düzeyde; ikisi ise, negatif yönlüdür. Dolayısıyla, her bir faktör kendi içinde benzeşiktir. Diğer faktörlerle karşılaştırıldığında ölçülen özellikler farklılaşsa da her bir faktörün TÖBT’ün ölçtüğü asıl psikolojik yapıya katkı sağladığı söylenebilir.

TÖBT’ün Ölçme Değişmezliği İncelemeleri: TÖBT’ün ülke ve cinsiyet değişkenleri açısından ölçme değişmezliğine sahip olup olmadığının belirlenmesinde EÖSA kullanılmıştır. Çözümlemeler, örtük sınıf modellerinin farklı homojenlik düzeylerinde incelenmiştir. Bu düzeylerden ilki, modelde yer alan tüm parametrelerin karşılaştırılan gruplar açısından serbest bırakıldığı heterojen modeldir. İkincisi, modele ilişkin çeşitli kısıtların koyulduğu kısmi homojen modeldir. Üçüncüsü ise, grup değişkenine ilişkin etkilerin (doğrudan ve ortak) olmadığı homojen modeldir. Buna göre ülke ve cinsiyet karşılaştırmalarında kullanılan modeller, heterojen modelden başlayarak homojen modele doğru hiyerarşik bir sırayla açıklanmıştır. TÖBT’ün ülkelere göre ölçme değişmezliğine sahip olup olmadığını belirlemek için yapılan kestirim değerleri, Tablo 8’de sunulmuştur.

Tablo 8. TÖBT'ün Ülkelere Göre Ölçme Değişmezliği Model Kestirimi

Model	LL	BIC (LL)	AIC (LL)	Par. Sayısı	L ²	sd	p-değeri	Sın. Hatası
Heterojen	-26490.33	56809.28	54072.67	546	39718.17	564	0.00	0.01
Kısmi Homojen	-26643.30	55432.31	53898.60	306	40024.11	804	0.00	0.01
Homojen	-26876.07	55028.35	54116.15	182	40489.66	928	0.00	0.02

Tablo 8'de farklı homojenlik düzeylerini gösteren modellerle, bu modellere ilişkin uygunluk ölçüleri, parametre sayısı, serbestlik derecesi ve L² istatistiğinin manidarlık değeri yer almaktadır. Model seçim sürecine geçmeden önce kurulan tüm modellere ilişkin serbestlik derecesine bakılmış ve tüm değerlerin pozitif olduğu görülmüştür. Bu durumda modelde yer alan parametre sayısı sütun sırasına eşittir ve modellerde tanımlanabilirlik sorunu bulunmamaktadır. Kurulan tüm modellere ait L² istatistiğinin manidarlık değeri, belirlenen manidarlık düzeyi sınırının altında kaldığından ($p < 0.05$) kurulan modelin uygun olmadığı yorumu yapılabilir. Ancak model değerlendirme sürecinde daha önce de belirtildiği gibi L² istatistiğinin, kay-kare dağılımını gerektirmesi ve örnekleme bağımlı bir yapıya sahip olması nedeniyle bilgi kriterlerinden yararlanılmıştır. Buna göre, BIC ve AIC değerleri incelendiğinde modelin homojenlik düzeyi arttıkça modellere ilişkin değerlerin (AIC değeri dışında) giderek azaldığı görülmektedir. Ancak AIC değeri, heterojen modelden kısmi homojen modele geçişte düşerken, homojen modele geçişte arttığı görülmektedir. Bu durum, kısmi homojen modelin uygun olduğunu işaret etmektedir. Bir başka deyişle, testin her üç ülkeye göre kısmen farklılaştığını göstermekte; testte yer alan maddelerin kültürden kısmen etkilendiğine ilişkin ipuçları vermektedir. Ancak yapılan araştırmaların örneklem büyüklüğünün 1000 civarında olduğu durumlarda BIC'in AIC'e göre daha tutarlı sonuçlar verdiğini göstermektedir (Dias, 2006; Kankaras, Vermunt ve Moors, 2011; Moors ve Wennekers, 2003; Morren, Gelissen ve Vermunt, 2011). Bu nedenle, model seçiminde formülasyonunda örneklem büyüklüğüne ilişkin düzeltme içeren BIC değeri dikkate alınmıştır. Bu değere göre veriye en uygun model homojen modeldir.

Bu belirlemeler ışığında TÖBT, Almanya, Azerbaycan ve İran açısından ölçme değişmezliğine sahiptir. Bu durum, TÖBT'ün her üç ülkede, aynı faktör yapısına sahip olduğu ve aynı psikolojik özellikleri ölçtüğü anlamına gelmektedir. Tablo 9'da homojen modele ilişkin bazı sınıflandırma istatistikleri verilmiştir.

Tablo 9. Ülkelere Göre Elde Edilen Homojen Modele İlişkin Sınıflandırma İstatistikleri

Sınıflandırma Hatası	0.02
Hata İndirgeme Değeri (λ)	0.97
Entropi R ² Değeri	0.96
Standart R ² Değeri	0.96

Tablo 9'da yer alan homojen modele ilişkin sınıflandırma hatasının oldukça düşük (%2) olduğu görülmektedir. Bu durumda modelin, %98 doğrulukla sınıflandırma yaptığı söylenebilir. Hata indirgeme, entropi R² ve standart R² değerlerinin sırasıyla, 0.97, 0.96 ve 0.96'dır. Dolayısıyla model kestirimi oldukça başarılıdır. Testin cinsiyete göre ölçme değişmezliğine sahip olup olmadığını gösteren model kestirimleri Tablo 10'da verilmiştir.

Tablo 10. TÖBT'ün Cinsiyete Göre Ölçme Değişmezliği Model Kestirimi

Model	LL	BIC (LL)	AIC (LL)	Par. Sayısı	L ²	sd	p-değeri	Sın. Hatası
Heterojen	-26752.05	56056.51	54232.10	364	39493.91	746	0.00	0.01
Kısmi Homojen	-26824.13	55359.21	54136.25	244	39638.06	866	0.00	0.01
Homojen	-26876.07	55028.35	54116.15	182	39741.96	928	0.00	0.02

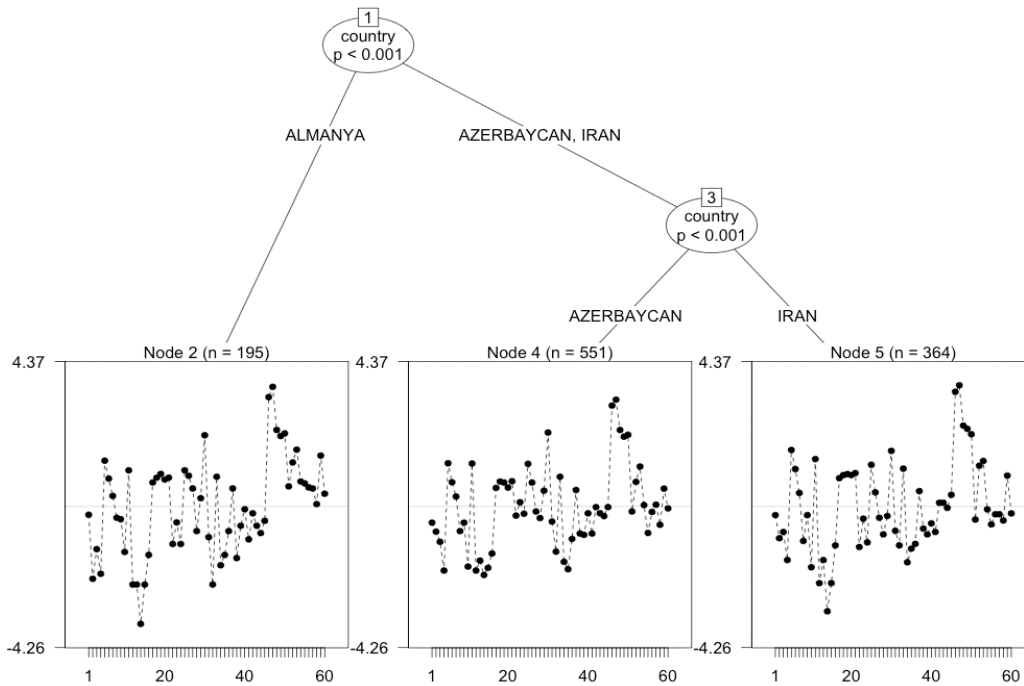
Tablo 10'da yer alan serbestlik derecelerinin tümünün pozitif olması, modellerde tanımlanabilirlik sorunun olmadığı anlamına gelmektedir. Modellere ait BIC ve AIC değerlerinin giderek azalması ise, TÖBT'ten elde edilen veriye en uygun modelin homojen model olduğunu göstermektedir. Bu belirlemeler ışığında TÖBT'ün, cinsiyet açısından ölçme değişmezliğine sahip olduğunu söylemek mümkündür. TÖBT'ün faktör yapısının hem kadın hem de erkek öğrenciler açısından aynı olması ve testin her iki cinsiyet açısından aynı psikolojik özellikleri ölçtüğü anlamına gelmektedir. Tablo 11'de, homojen modele ilişkin bazı sınıflandırma istatistikleri verilmiştir.

Tablo 11. Cinsiyete Göre Elde Edilen Homojen Modele İlişkin Sınıflandırma İstatistikleri

Sınıflandırma Hatası	0.02
Hata İndirgeme Değeri (λ)	0.97
Entropi R ² Değeri	0.96
Standart R ² Değeri	0.96

Tablo 11'e göre sınıflandırma hatası oldukça düşük (%2)'tür. Dolayısıyla modelin %98 doğrulukla sınıflandırma yaptığı söylenebilir. Hata indirgeme, entropi R² ve standart R² değerleri ise sırasıyla, 0.97, 0.96 ve 0.96'dır. Bu değerler, model kestiriminin oldukça başarılı olduğunu göstermektedir.

TÖBT'ün Madde İşlev Farklılığı İncelemeleri: TÖBT'te yer alan her bir maddenin ülkelere ve cinsiyete göre MİF gösterip göstermediği Rasch ağacı oluşturularak belirlenmiştir. Ancak ÖBA'nın bu ağacı oluşturabilmesi için bir başlangıç değişkenine ihtiyacı vardır. MİF'in iki ya da daha fazla değişken açısından birlikte ele alındığı çalışmalar için bu başlangıç noktası, her bir değişkene ait manidarlık değeridir. Her bir değişken için hesaplanan manidarlık değeri ne kadar küçükse, R programı Rasch ağacını oluşturabilmek için bölümlenmeye o değişkenden başlamaktadır (Zeileis vd., 2008). Bu araştırmada MİF, ülke ve cinsiyet değişkenleri açısından ayrı ayrı ele alınarak incelenmiştir. Testte yer alan maddelerin ülkelere göre MİF gösterip göstermediğini belirleyebilmek için maddelere verilen yanıtların yer aldığı gözlenen değişkenler matrisine dayalı olarak oluşturulan ve amacı bu matrisi homojen alt gruplara ayırmak olan Rasch ağacı Şekil 1'de sunulmuştur.



Şekil 1. Ünelere Göre Rasch Ağacı⁴

⁴ Country: Ülke, Node: Boğum

Şekil 1’de yer alan Rasch ağacındaki en son boğumlar (nodes), TÖBT’te yer alan 60 maddeye ilişkin ülkelere göre elde edilen madde parametre kestirimlerini göstermektedir. Buna göre, madde güçlük parametrelerine ilişkin kestirim değerleri 4.37 ile -4.26 arasındadır. Bu değerlerin yüksekliği maddelerin güç, düşüklüğü ise kolay olduğu anlamına gelmektedir (Strobl vd., 2015). Ayrıca şekildeki her bir bölünmeye ilişkin manidarlık değerleri de yer almaktadır.

Bu bilgiler ışığında bölünmenin gerçekleştiği ülkeler açısından manidar düzeyde ($p<0.001$) bir farklılaşma görülmektedir. Elde edilen Rasch ağacı, farklılığın hangi ülkelerde olduğuna ilişkin bilgiyi de sunmaktadır. Buna göre ilk bölünmede Almanya, diğer iki ülkeye (Azerbaycan ve İran) göre, ikinci bölünmede ise Azerbaycan, İran’a göre farklılaşmaktadır. Rasch ağacı ayrıntılı olarak incelendiğinde, TÖBT’te yer alan bazı maddeler ülkeler açısından MİF içermektedir. Buna göre, MİF gösteren maddeler ülkeler açısından karşılaştırmalı olarak sunulmuştur.

23. *madde*; parça-bütün ilişkisi kurmayı gerektiren Faktör 1’in altında yer almaktadır. Bu madde, Alman ve İranlı bireyler için kolay, Azerbaycanlı bireyler için bir miktar güç görünmektedir. Madde, Azerbaycanlı bireyler lehine bir yanlılık oluşturuyor gibi görünse de maddelerin ortalama güçlük düzeyi dikkate alındığında, bu maddenin güçlük düzeyi her üç ülke için de sifıra yakındır.

27. *madde*; analitik düşünmeyi gerektiren Faktör 3’ün altında yer almaktadır. Bu madde, Azerbaycan ve İranlı bireyler için kolay, Alman bireyler için güç görünmektedir. Madde, Azerbaycanlı ve İranlı bireyler lehine bir yanlılık oluşturuyor gibi görünse de maddelerin ortalama güçlük düzeyi dikkate alındığında, bu maddenin güçlük düzeyi her üç ülke için de sifıra yakındır.

29. *madde*; çıkarım yapmayı gerektiren Faktör 2’nin altında yer almaktadır. Bu madde, Alman ve Azerbaycanlı bireyler için güç, İranlı bireyler için kolay görünmektedir. Madde, İranlı bireyler lehine bir yanlılık oluşturuyor gibi görünse de maddeler her üç ülkenin bireyleri açısından ortalama güçlük düzeyine çok yakındır.

42-45 *arası maddeler*; analitik düşünmeyi gerektiren Faktör 3’ün altında yer almaktadır. Bu maddeler, Alman ve Azerbaycanlı bireyler için kolay, İranlı bireyler için güç görünmektedir. Maddeler, özellikle Alman ve bir miktar da Azerbaycanlı bireyler lehine bir yanlılık oluşturmaktadır. Ancak Azerbaycanlı ve İranlı bireyler açısından maddeler ortalama güçlük düzeyine çok yakındır.

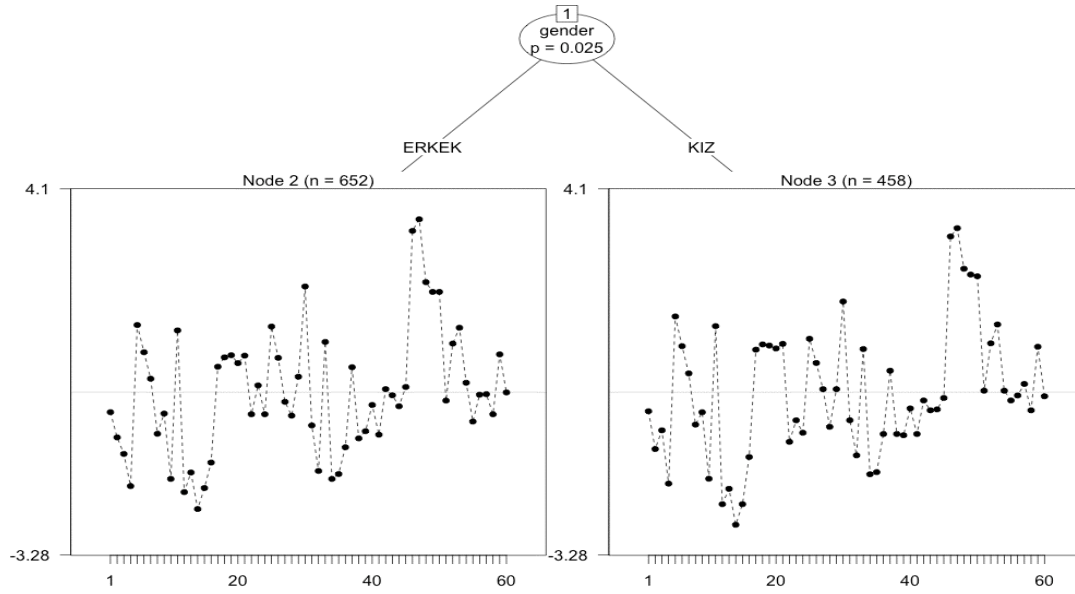
51. *madde*; analitik düşünmeyi gerektiren Faktör 3’ün altında yer almaktadır. Bu madde, Azerbaycan ve İranlı bireyler için kolay, Alman bireyler için güç görünmektedir. Madde, Azerbaycanlı ve İranlı bireyler lehine bir yanlılık oluşturuyor gibi görünse de madde ortalama güçlük düzeyindedir.

54-58 *arası maddeler*; analitik düşünmeyi gerektiren Faktör 3’ün altında yer almaktadır. Bu maddeler, Azerbaycanlı ve İranlı bireyler için kolay, Alman bireyler için güç görünmektedir. Bu grupta yer alan maddeler, Azerbaycanlı ve İranlı bireyler lehine bir yanlılık oluşturuyor gibi görünse de maddelerin ortalama güçlük düzeyine yakın gözden uzak tutulmamalıdır.

60. *madde*; analitik düşünmeyi gerektiren Faktör 3’ün altında yer almaktadır. Bu madde, Azerbaycan ve İranlı bireyler için bir miktar kolay, Alman bireyler için bir miktar güç görünmektedir. Madde, Azerbaycanlı ve İranlı bireyler lehine bir yanlılık oluşturuyor gibi görünse de maddenin ortalama güçlük düzeyinde olduğu görülmektedir.

MİF gösteren 14 maddeye bir bütün olarak bakıldığından, maddelerin ülkelere göre güçlük ve kolaylık düzeylerinin ortalama değer olan sıfır değerine yakın olduğu dikkat çekmektedir. Bu durum ülkeler açısından maddelerde belirgin bir yanlılık olmadığını göstermektedir. Bunun dışında kalan 46 maddenin ise birbirilerine yakın dağılıma ve değerlere sahip olduğu görülmektedir.

Testte yer alan maddelerin cinsiyete göre MİF gösterip göstermediğini belirleyebilmek için maddelere verilen yanıtların yer aldığı gözlenen değişkenler matrisine dayalı olarak oluşturulan ve amacı bu matrisi homojen alt gruplara ayırmak olan Rasch ağacı Şekil 2’de verilmiştir.



Şekil 2. Cinsiyete Göre Rasch Ağacı⁵

Şekil 2’de yer alan Rasch ağacındaki boğumlar, TÖBT’te yer alan 60 maddeye ilişkin cinsiyete göre elde edilen madde parametre kestirimlerini göstermektedir. Buna göre madde güçlük parametrelerine ilişkin kestirim değerleri, 4.10 ile -3.28 arasındadır. Bu değerlerin yüksek ve düşük olduğu durumlar için yapılacak yorumlar ülkelere ilişkin yorumlardaki gibidir.

Bu bilgiler ışığında bölünmenin gerçekleştiği kadınlar ve erkekler açısından manidar düzeyde ($p>0.01$) bir farklılaşma görülmektedir. Elde edilen Rasch ağacı, TÖBT’te yer alan bazı maddelerin (örneğin 23, 42, 45, 51 ve 57) cinsiyet açısından MİF gösterdiği yönünde bir bilgi verse de, bölümlenmeye ilişkin p değerinin manidar olmaması, maddelerin cinsiyet açısından MİF içermediğini göstermektedir.

Apinyapibal, Lawthong ve Kanjanawasee (2015), ÖBA ile elde edilen Rasch ağacı sayesinde hem önceden tanımlanmış hem de örtük grupların MİF açısından tespit edilebildiği; yöntemin çıktısı olan ve kolaylıkla anlaşılacak bir diyagram şeklinde gösterilen sonuçlar sayesinde madde gruplarının açıklanabildiğini vurgulamışlardır.

Tartışma ve Sonuç

Araştırmanın sonuçlarına bir bütün olarak bakıldığında ilk olarak, AYÖS 2013’te kullanılan TÖBT’te yer alan 60 maddenin üç faktörlü bir yapıdan oluştuğu görülmektedir. Testte yer alan maddelerin faktörlere dağılımına bakıldığında; parça-bütün ilişkisi kurma faktöründe 12 madde, çıkarım yapma faktöründe 14 madde ve analitik düşünme faktöründe 34 madde yer almaktadır. Üçüncü faktörde diğer faktörlere göre daha fazla sayıda madde toplanmasının nedeni, bu faktörün birleştirme, ayırıştırma ve parçalar arasında ilişkileri bulmayı gerektiren maddelerden oluşması olabilir. Analitik düşünme, bütünü anlamak amacıyla kendisini oluşturan parçaları zihinde soyut olarak ayırma ve parçalar arasındaki ilişkileri inceleme süreci olarak tanımlanmaktadır (Bruner, 1957, aktaran Lohman ve Lakin, 2011). TÖBT’ün ölçmeye çalıştığı yapı bir bütün olarak düşünüldüğünde test, şekil, sayı ve harf ilişkileri üzerine kuruludur. Testte yer alan maddeler bu özellikler arasındaki ilişkilere göre kurgulanmaktadır. Bu yönüyle TÖBT, uluslararası düzeyde aynı amaçlarla geliştirilen örnek testlere benzer bir yapı sergilemektedir. Bu tür testler üzerinde yapılan yapı geçerliği çalışmaları testlerin analitik düşünme, eleştirel düşünme, sonuç çıkarma, problem çözme, akıl yürütme ve tümevarım gibi bilişsel süreçlerden oluştuğunu göstermektedir (Enright ve Powers, 1991; Lord ve Wild, 1985; McCallum, 2003).

⁵Gender: Cinsiyet, Node: Boğum

Testin faktör yapısının belirlenmesinde ÖSFA kullanılmıştır. Bu analizin kullanılma nedeni, verilerin elde edildiği testin yapısının sıralama düzeyinde olması ve testte yer alan maddelerin ikili yanıtlardan oluşmasıdır. İki kategorili yanıtlardan elde edilen dikotom (1-0, var-yok, doğru-yanlış gibi) veri yapısına sahip çalışmalarda genel uygulama, FA'da kullanılan teknikleri tetrakorik korelasyon matrisi üzerinden yapmaktır. Ancak bu yöntem birçok gerekliliğin yanı sıra Gramian/Gram Korelasyon Matrisini gerektirmekte ve genellikle sonuçlarda bozulmaya yol açmaktadır (Christofferson, 1975). Yalnızca sürekli değişkenler için uygun olan FA'nın uygulamada sıklıkla, sınıflama ve sıralama düzeyindeki değişkenleri içeren ölçek türlerinde de kullanıldığı görülmektedir. Bu durum, birçok açıdan doğrusallık varsayımının ihlal edilmesine ve gerçek model özünde doğrusal değilken doğrusal bir modelmiş gibi yorumlanmasına neden olmaktadır (Magidson ve Vermunt, 2003b). ÖSFA yaklaşımı veri yapısını açıklamada ve yorumlamada daha kolay bir yaklaşım sunması açısından da tercih edilmelidir. Bunun yanı sıra çok boyutluluk söz konusu olduğunda ÖSFA modellerinin kullanılması daha uygundur (Moors, 2003). Dolayısıyla bu çalışmada FA'nın kategorik değişkenler için kullanılması uygun görülmemiş; bunun yerine örtük sınıf modelleri gibi kategorik veri yapısına uygun olan boyut indirgeme yolları tercih edilmiştir.

Testin ülkelere ve cinsiyete göre ölçme değişmezliğine sahip olup olmadığı EÖSA yardımıyla, heterojen modelden başlayarak homojen modele doğru hiyerarşik bir sırayla ve farklı homojenlik düzeylerinde incelenmiştir. Her iki değişken açısından yapılan ölçme değişmezliği incelemelerinde homojen modelin seçilmiş olması, TÖBT'ün hem ülke hem de cinsiyet açısından ölçme değişmezliğine sahip olduğuna işaret etmektedir. Bu durum testin üç faktörlü yapısının hem ülkelere hem de cinsiyete göre aynı olduğunu; bir başka deyişle, testin her iki değişken açısından aynı psikolojik özellikleri ölçtüğünü göstermektedir. Bu beklenen bir durumdur; çünkü TÖBT hem kullanım amacı olarak hem de bu testi alan farklı kültürlerden gelen öğrenciler düşünüldüğünde, dilden bağımsız bir ölçme yapmayı hedeflemektedir. Oldukça farklı birey profillerine sahip farklı kültürler düşünüldüğünde, bu eşdeğerliği sağlamak zor olsa da alanyazında bu tür testlerin kültürel özelliklerden bağımsız ölçme yapabileceğini vurgulayan çalışmalar yer almaktadır (Cattell, 1979; Raven, 2000).

Testin ölçme değişmezliğinin incelenmesinde de faktör analizine dayalı yaklaşımların yerine örtük sınıf analizine dayalı yaklaşımlar tercih edilmiştir. Ölçme değişmezliğinin incelenmesinde kullanılan istatistiksel yöntemler yaygın olarak ya yapısal eşitlik modelini ya da ortalama ve kovaryans modelini temel alan ÇGDFA'ya dayanmaktadır (Jöreskog, 1971; Sörbom, 1974). Ancak bu türden yöntemler, faktör analitik yöntemler gibi kullandıkları korelasyon matrisinin türü ve normal dağılım varsayımı gerektirmeleri gibi nedenlerle kategorik veri yapısını çözümlenmede her zaman uygun bir yaklaşım sunamamaktadır. Oysa örtük sınıf modeli temelli analizler, oldukça esnek bir yaklaşım sunarak her türlü ölçme yanlılıklarının belirlenmesinde sağlam kanıtlar ortaya koymaktadır (Moors, 2003).

Araştırmada ölçme değişmezliği incelemelerinde kullanılan model seçimi kriteri olarak ülkeler açısından yapılan karşılaştırmalarda, formülasyonunda örneklem büyüklüğüne ilişkin düzeltme içeren BIC değeri dikkate alınmıştır. Bunun nedeni BIC ve CAIC'in küçük örneklemelerde kullanılmasının uygun olmamasıdır (Güngör Culha, 2012). Bir diğer nedeni de BIC değerinin daha sade modeli seçme eğiliminde olmasıdır. Bu sayede kestirim değerleri daha kolay yorumlanabilmektedir. Ayrıca, yapılan araştırmalar örneklem büyüklüğünün 1000 civarında olduğu durumlarda BIC'in daha tutarlı sonuçlar verdiğini belirtmektedir (Dias, 2006; Kankaras, Vermunt ve Moors, 2011; Moors ve Wennekers, 2003; Morren vd., 2011). Model seçiminde alternatif bir yol olarak AIC ve BIC değerlerinin kullanılması genel kabul görmüş olmakla birlikte, AIC3, CAIC, SABIC gibi türevleri de var olan kriterlerden hangisinin kullanılmasının uygun olduğuna dair bir görüş birliği bulunmamaktadır. Bu konuyla ilgili alanyazında birçok çalışma yer almaktadır (Bauer ve Curran, 2003; Dias, 2006; Lin, 2006; Nylund, Asparouhov ve Muthen, 2007; Vrieze, 2012; Yang ve Yang, 2007, aktaran Güngör, Korkmaz ve Sazak, 2015).

Testte yer alan maddelerin ülkelere ve cinsiyete göre madde işlev farklılığı gösterip göstermediği Rasch modeline dayalı Özyinelemeli Bölümleme Analiziyle incelenmiştir. Her iki değişken açısından yapılan madde işlev farklılığı incelemeleri sonucunda, 14 maddenin ülke değişkeni

açısından farklı işlev gösterdiği görülmüştür. MİF gösteren 14 maddeye bir bütün olarak bakıldığında, maddelerin ülkelere göre güçlük ve kolaylık düzeylerinin ortalama değer olan sıfır değerine yakın olduğu dikkat çekmektedir. Bunun dışında kalan 46 madde birbirine yakın dağılıma ve değerlere sahiptir.

Westers ve Kelderman (1992), eğitimde ve psikolojide kullanılan ölçme araçlarının manidar düzeyde MİF gösteren maddeler içeriyor olmasının bazı alt gruplar açısından adil olmayacağını vurgulayarak; MİF gösteren maddelerin saptanarak bu türden maddelerin revize edilmesi ya da ölçme aracından tamamen çıkarılması gerektiğini belirtmektedirler. Ancak MİF belirlenen maddelerin doğrudan testten çıkarılması özellikle az sayıda maddenin bulunduğu ya da ölçülen psikolojik yapıya ilişkin kapsam geçerliğinin zarar göreceği durumlarda sorun yaratabilmektedir. Testte yer alan maddeler böylesi durumlar göz önünde bulundurularak deneysel yollarla saptandıktan sonra ilgili maddelerin testi alan alt gruplar için avantaj ya da dezavantaj sağlayıp sağlamadığı uzmanlar tarafından da ortaya koyulmalıdır (Hambleton, Swaminathan ve Rogers, 1991).

Cinsiyete göre yapılan incelemelerde ise, kadınlar ve erkekler açısından manidar düzeyde bir farklılaşma olmadığı; bir başka deyişle testte yer alan maddelerin cinsiyet açısından MİF içermediği sonucuna ulaşılmıştır. Cinsiyete göre yapılan incelemelerde ise, kadınlar ve erkekler açısından manidar düzeyde bir farklılaşma olmadığı; bir başka deyişle testte yer alan maddelerin cinsiyet açısından MİF içermediği sonucuna ulaşılmıştır.

Öneriler

Bu araştırmada, TÖBT'ün ülkelere ve cinsiyete göre ölçme değişmezliğine sahip olup olmadığı; testte yer alan maddelerin bu değişkenlere göre farklı işlev gösterip göstermediği nicel çözümleme yolları kullanılarak incelenmiştir. Buna göre MİF gösteren maddelere dayalı olarak, maddeyi yanıtlayan bireyler ile madde yazarları üzerinde yapılacak çalışmalarla testte yer alan maddelerin farklı işlev gösterme nedenleri kapsamlı olarak belirlenebilir.

Bu araştırma kapsamında yapılan değişmezlik ve madde işlev farklılığı incelemeleri yalnızca ülke ve cinsiyet değişkenleri açısından yapılmıştır. Başka araştırmalarla yaş, okul türü, bölge, azınlık olma durumu vb. değişkenler de ele alınıp incelenebilir. Bunun yanı sıra araştırma Almanya, Azerbaycan ve İran'dan AYÖS'e katılan bireylerin TÖBT'e verdikleri yanıtlar üzerinden gerçekleştirilmiştir. TÖBT, ilgili ülkelerde AYÖS'e giren bireylerle benzer özelliklere sahip bireylerden oluşan gruplara uygulanarak karşılaştırmalı çalışmalar yapılabilir.

Son olarak bu araştırma, bir test uygulamasından sonra elde edilen yanıt örüntüleri kullanılarak yapılan kestirimlere dayalı olarak kurgulanmıştır. Başka araştırmalarla incelenecek testler, asıl uygulamaları yapılmadan önce benzer gruplar üzerinde denenebilir ve geçerlik için önemli olan yansız testler elde edilebilir.

Teşekkür

Bu çalışmanın verileri ANKÜDEM'den alınmıştır.

Kaynakça

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington: American Educational Research Association.
- ANKÜDEM. (2011). *Ankara Üniversitesi Yabancı Uyumlu Öğrenci Seçme ve Yerleştirme Sınavı (AYÖS) projesi kesin raporu*. Proje No: 11Y5250001. Ankara: Ankara Üniversitesi Bilimsel Araştırma Projeleri Ofisi.
- ANKÜDEM. (2012). *AYÖS 2012 Temel Öğrenme Becerileri Testi üzerine bir çalışma*. Kurum içi Rapor. Ankara: Ankara Üniversitesi Ölçme ve Değerlendirme Uygulama ve Araştırma Merkezi.
- Apinyapibal, S., Lawthong, N. ve Kanjanawasee, S. (2015). A comparative analysis of the efficacy of differential item functioning detection for dichotomously scored items among logistic regression, SIBTEST and Rasch tree methods. *Procedia-Social and Behavioral Sciences*, 191(2), 21-25.
- Bartholomew, D. J., Steele, F., Moustaki, I. ve Galbraith, J. I. (2008). *Analysis of multivariate social science data*. (2. bs.). Boca Raton: Taylor and Francis Group.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley-Interscience Publication.
- Cattell, R. B. (1979). Are culture fair intelligence tests possible and necessary? *Journal of Research and Development in Education*, 12(2), 3-13.
- Clogg, C. C. ve Goodman, L. A. (1984). Latent structure analysis of a set of multidimensional contingency tables. *Journal of the American Statistical Association*, 79(388), 762-771.
- Clogg, C. C. ve Goodman, L. A. (1985). Simultaneous latent structure analysis in several groups. *Sociological Methodology*, 15, 81-110.
- Collins, L. M. ve Lanza, S. T. (2010). *Latent class and latent transition analysis with application in the social, behavioral, & health sciences*. New Jersey: John Wiley and Sons, Inc.
- Christofferson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika*, 40(1), 5-32.
- Dias, J. G. (2006). Latent class analysis and model selection. M. Spiliopoulou, R. Kruse, C. Borgelt, A. Nurnberger ve W. Gaul (Ed.), *From data and information analysis to knowledge engineering* içinde (s. 95-102). Berlin: Springer Heidelberg.
- Enright, M. K. ve Powers, D. E. (1991). *Validating the GRE analytical ability measure against faculty ratings of analytical reasoning skills*. New Jersey: Lawrence Erlbaum Associates, Inc.
- Guilford, J. P. (1954). *Psychometric methods* (2. bs.). New York: McGraw-Hill.
- Güngör Culha, D. (2012). *Örtük sınıf analizlerinde ölçme eşdeğerliğinin* (Yayımlanmamış doktora tezi). Ege Üniversitesi, İzmir.
- Güngör D., Korkmaz, M. ve Sazak, H. S. (2015). Örtük sınıf analiziyle yapılan ölçme eşdeğerliği çalışmalarında model seçimi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 30(1), 90-105.
- Hagenaars, J. A. ve McCutcheon, A. L. (Ed.). (2002). *Applied latent class analysis*. Cambridge: Cambridge University Press.
- Hambleton, R. K., Swaminathan, H. ve Rogers, H. J. (1991). *Fundamentals of item response theory*. California: Sage Publications, Inc.
- Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36, 409-426.
- Kankaras, M. ve Moors, G. B. D. (2009). Measurement equivalence in solidarity attitudes in Europe: Insights from a multiple-group latent-class factor approach. *International Sociology*, 24, 557-579.
- Kankaras, M. ve Vermunt, J. K. (2014). Simultaneous latent class analysis across groups. A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* içinde (s. 5969-5974). Dordrecht: Springer.
- Kankaras, M., Moors, G. B. D. ve Vermunt, J. K. (2011). Testing for measurement invariance with latent class analysis. E. Davidov, P. Schmidt ve J. Billiet (Ed.), *Cross-cultural analysis: Methods and applications* içinde (s. 359-384). New York: Taylor and Francis Group.

- Kankaras, M., Vermunt, J. K. ve Moors, G. B. D. (2011). Measurement equivalence of ordinal items. A comparison of factor analytic, item response theory, & latent class approaches. *Sociological Methods and Research*, 40(2), 279-310.
- Kopf, J., Augustin, T. ve Strobl, C. (2010). *The potential of model-based recursive partitioning in the social sciences: Revisiting Ockham's Razor* (Technical report number 88). Munich: University of Munich, Department of Statistics.
- Kuder, G. F. ve Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2(3), 151-160.
- Lin, H. T. (2006). A comparison of model selection indices for nested latent class models. *Monte Carlo Methods and Applications*, 12(3), 239-259.
- Lohman, D. F. ve Lakin, J. M. (2011). Reasoning and intelligence. R. J. Sternberg ve S. B. Kaufman (Ed.), *Handbook of intelligence* içinde (s. 419-441). New York: Cambridge University Press.
- Lord, F. M. ve Wild, C. L. (1985). Contribution of verbal item types in the GRE general test to accuracy of measurement of the verbal scores. GRE Board Professional Report GREB No. 84-6P, ETS Research Report No. 85-29. New Jersey: Educational Testing Service. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.495.1909&rep=rep1&type=pdf> adresinden erişildi.
- Magidson, J. ve Vermunt, J. K. (2001). Latent class factor and cluster models, bi-plots and related graphical displays. *Sociological Methodology*, 31(1), 223-264.
- Magidson, J. ve Vermunt, J. K. (2003a). *A nontechnical introduction to latent class models*. White Paper. Statistical Innovations.
- Magidson, J. ve Vermunt, J. K. (2003b). Comparing latent class factor analysis with the traditional approach in data mining. H. Bozdoğan (Ed.), *Statistical data mining and knowledge discovery* içinde (s. 373-383). Boca Raton: Chapman and Hall/CRC.
- Magidson, J. ve Vermunt, J. K. (2004). Latent class models. D. Kaplan (Ed.), *The sage handbook of quantitative methodology for the social sciences* içinde (s. 175-198). Thousand Oaks: Sage Publications, Inc.
- McCallum, R. S. (2003). Context for nonverbal assessment of intelligence and related abilities. R. S. McCallum (Ed.), *Handbook of nonverbal assessment* içinde (s. 3-18). New York: Kluwer Academic/Plenum Publishers.
- McCutcheon, A. L. (1987). *Latent class analysis*. Sage University Papers Series. Quantitative Applications in the Social Sciences. No. 07-064. Newbury Park: Sage Publications, Inc.
- McCutcheon, A. L. (2002). Basic concepts and procedures in single- and multiple-group latent class analysis. J. A. Hagenaars ve A. L. McCutcheon (Ed.), *Applied latent class analysis* içinde (s. 56-86). Cambridge: Cambridge University Press.
- Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research: Applications of Item Response Theory*, 13, 123-144.
- Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*, 58, 525-543.
- Messick, S. (1989). Validity. R. L. Linn (Ed.), *Educational measurement* (3. bs., s. 13-103). New Jersey: American Council on Education and Macmillan Publishing Company.
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York: Taylor and Francis Group.
- Moors, G. B. D. (2003). Diagnosing response style behavior by means of a latent-class factor approach. Socio-demographic correlates of gender role attitudes and perceptions of ethnic discrimination reexamined. *Quality and Quantity*, 37(3), 277-302.

- Moors, G. B. D. ve Wennekers, C. H. W. (2003). Comparing moral values in western European countries between 1981 and 1999. A multigroup latent-class factor approach. *International Journal of Comparative Sociology*, 44(2), 155-172.
- Morren, M., Gelissen, J. ve Vermunt, J. K. (2011). Dealing with extreme response style in cross-cultural research: A restricted latent class factor analysis approach. *Sociological Methodology*, 41(1), 13-47.
- Raven, J. (2000). The Raven's progressive matrices: Change and stability over culture and time. *Cognitive Psychology*, 41(1), 1-48.
- Shealy, R. ve Stout, W. F. (1993). A model-based standardization approach that separates true bias/DIF from group differences and detects test bias/DTF as well as item bias/DIF. *Psychometrika*, 58, 159-194.
- Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology*, 27, 229-239.
- Strobl, C., Kopf, J. ve Zeileis, A. (2015). A new method for detecting differential item functioning in the Rasch model. *Psychometrika*, 80(2), 289-316.
- Strobl, C., Wickelmaier, F. ve Zeileis, A. (2011). Accounting for individual differences in Bradley-Terry models by means of recursive partitioning. *Journal of Educational and Behavioral Statistics*, 36(2), 135-153.
- Vandenberg, R. J. ve Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, & recommendations for organizational research. *Organizational Research Methods*, 3, 4-70.
- Vermunt, J. K. ve Magidson, J. (2000). Graphical displays for latent class cluster and latent class factor models. W. Jansen ve J. G. Bethlehem (Ed.), *Proceedings in computational statistics 2000* içinde (s. 121-122).
- Vermunt, J. K. ve Magidson, J. (2005a). Factor analysis with categorical indicators: A comparison between traditional and latent class approaches. L. A. Van der Ark, M. A. Croon ve K. Sijtsma (Ed.), *New developments in categorical data analysis for the social and behavioral sciences* içinde (s. 41-62). Mahwah: Lawrence Erlbaum Associates, Inc.
- Vermunt, J. K. ve Magidson, J. (2005b). *Latent GOLD 4.0 user's guide*. Massachusetts: Statistical Innovations Inc.
- Vermunt, J. K. ve Magidson, J. (2013a). *Latent GOLD 5.0 upgrade manual*. Belmont: Statistical Innovations Inc.
- Vermunt, J. K. ve Magidson, J. (2013b). *Technical guide for Latent GOLD 5.0: Basic, advanced, & syntax*. Belmont: Statistical Innovations Inc.
- Vermunt, J. K. ve Magidson, J. (2013c). *LG-Syntax user's guide: Manual for Latent GOLD 5.0 syntax module*. Belmont: Statistical Innovations Inc.
- Westers, P. ve Kelderman, H. (1992). Examining differential item functioning due to item difficulty and alternative attractiveness. *Psychometrika*, 57(1), 107-118.
- Wolcott, H. F. (1994). *Transforming qualitative data: Description, analysis, & interpretation*. Thousand Oaks, California: Sage Publications, Inc.
- YÖK. (2013). *Yurtdışından öğrenci kabulüne ilişkin esaslar*. Karar Tarihi: 01.02.2013. http://www.yok.gov.tr/documents/10279/19575784/Yurtdisindan_Ogrenci_Kabulune_Iliskin_Esaslara161115.pdf adresinden erişildi.
- Zeileis, A., Hothorn, T. ve Hornik, K. (2008). Model-based recursive partitioning. *Journal of Computational and Graphical Statistics*, 17(2), 492-514.
- Zeileis, A., Strobl, C., Wickelmaier, F. ve Kopf, J. (2011). *Psychotree: Recursive partitioning based on psychometric models*. R package version 0.12-1. <http://CRAN.R-project.org/package=psychotree> adresinden erişildi.